ANNI JÜRINE

The Development of Complex
Postpositions in Estonian:
A Case of Grammaticalization
via Lexicalization

DISSERTATIONES PHILOLOGIAE ESTONICAE
UNIVERSITATIS TARTUENSIS

38

ANNI JÜRINE

The Development of Complex Postpositions in Estonian: A Case of Grammaticalization via Lexicalization

University of Tartu, Institute of Estonian and General Linguistics

Dissertation accepted for the commencement of the degree of Doctor of Philosophy on November 19th, 2015 by the Committee of the Institute of Estonian and General Linguistics, Faculty of Philosophy, University of Tartu

Supervisors: Associate Professor Dr. Külli Habicht, University of Tartu

Associate Professor Dr. Ilona Tragel, University of Tartu

Opponent: Prof. Dr. Hubert Cuyckens, University of Leuven

Commencement: February 16th, 2015 at 14.15, Senat Hall in the University main building, Ülikooli 18, Tartu

European Union
European Social Fund

Investing in your future

University of Tartu Press
www.tyk.ee

PREFACE

I had my first predicament (that I can recall) with postpositional semantics when I was 3 or 4 years old. As I know now, I had reanalyzed a lexicalized postpositional phrase *töö juures* ('at workplace' literally 'work root-in)' and ended up with an unacceptable nominal compound *tööjuur* 'worplace', literally 'work root' (*ema töö juures* > *ema tööjuure*-s > *ema tööjuur*). Of course, kids make this kind of 'mistakes' all the time – they negotiate over meaning and language structure; that is how they acquire language. The reason I remember this incident so vividly is probably because everybody laughed at me and teased me about my misinterpretation. But perhaps also because I made an observation that the gram *juures* ('at') is connected to the noun referring to root (*juur*), which – as I like to think now – triggered the fascination with the peculiar ways in which language works. Although this fascination followed me throughout my schoolyears, my path to linguistics was anfractuous and full of lucky coincidences. Unsure of which specialty to choose, I followed my secondary school desk mate Kairit to University of Tartu to study Estonian and Finno-Ugric Linguistics. Determined to change my major and transfer to another discipline as soon as possible, I became a student of Estonian language. However, as soon as I had – again by coincidence – stumbled upon the topic of complex postpositions, I knew that I was not done with linguistics yet.

Completing my dissertation has taken a bit longer than I first anticipated when I started my doctoral studies in 2009. The angle on the topic has changed quite a bit compared to what it was when I started my investigation of the complex postpositions in Estonian. Even though I always knew that I wanted to write a monograph, I was unsure of how this endeavor would turn out until the very last minute. It is beyond certain that I would not have completed my dissertation without the ever present support of my supervisors Külli Habicht and Ilona Tragel. I am truly thankful to Külli, who has been my supervisor since my bachelor studies, for handing me this topic and letting me find my own way with it. I am indebted to her for guiding me throughout this whole process. Her door has always been open for me to drop by to discuss complex postpositions or whatever else was necessary. Our discussions have been a great inspiration to me. It goes without saying that I am as grateful to Ilona who joined our complex postpositions team in 2009 and offered us a fresh angle on the topic. She, too, has offered me guidance, encouragement and in many respects her support has gone beyond the 'job description' of any supervisor. You have been extremely patient and supportive of me, and for that I am truly grateful.

I am also thankful to my colleagues and fellow students at the Institute of Estonian and General Linguistics. I am especially grateful to Piia Taremaa for valuable comments on the manuscript of my thesis, Ann Veismann for fruitful discussions and feedback on my research over the years, and Helle Metslang for comments on the manuscript, support and encouragement. I am thankful to Jane Klavan for showing me the ropes of academic life, but especially for good com-

CONTENTS

I INTRODUCTION

1.1. The object and aims of the study

The present study concerns the development of Estonian complex postpositions and complex adverbs. This phenomenon is described as an instance of ongoing grammaticalization whereby simple postpositional phrases (as in (1)) become complex units that may function as complex adverbs (2) or as complex postpositions (3). From here on, I will refer to both parts of speech as complex function words.

(1) *See   tädi   lenda-s   pikali   ja   ma   vääna-si-n   ta-Ø*
    this   lady   fly-PST.3SG   down   and   I   bend-PST-1SG   s/he-GEN
    *käe-d   **selja-Ø   taha.***
    hand-PL   back-GEN   behind.LAT
    'This lady fell down and I bent her hands behind her back.' [filmitalgud.ee]

(2) *Subaru-ga   trikitamise-Ø   **selja-Ø   taha***   *jät-nud*
    Subaru-COM   maneuvering-GEN   back-GEN   behind.LAT   leave-PST.PTCP
    autotaltsutaja   Ken   Block   on   esimene   Ameerika-Ø
    race car driver   Ken   Block   be.3SG   first   America-GEN
    Ühendriiki-de   kodanik, kes   WRC-s   tantsu-Ø   löö-ma
    united states-PL.GEN   citizen who   WRC-INE   dance-PRT   hit-SUP
    hakka-b.
    start-3SG
    Lit. Race car driver Ken Block who has left doing stunts with Subaru behind his back is the first American citizen to join the WRC'
    'Race car driver Ken Block, whose days doing stunts with Subaru are over, is the first American citizen to join the WRC.' [www.delfi.ee]

(3) *Eestlas-te   **selja-Ø   taha***   *jä-i   ligemale*
    Estonian-PL.GEN   back-GEN   behind.LAT   stay-PST.3SG   around
    nelja-Ø   sekundi-Ø   kauguse-le   Itaalia   ja
    four-GEN   second-GEN   distance-ALL Italy   and
    viimase-na   lõpeta-s   Ungari-Ø   tandem.
    last-ESS   finish-PST.3SG   Hungary-GEN   tandem
    Lit. Italy stayed behind the back of the Estonians about four seconds away and Hungarian tandem finished last.
    'Italy came in about four seconds behind the Estonians and the Hungarian tandem finished last.' [www.sakala.ajaleht.ee]

The instance of grammaticalization investigated here involves a category shift – forms that are traditionally analyzed as freely combined phrases are now becoming holistically analyzed units that may function as adverbs or postpositions. Notably, all of the above usages are present in the contemporary Estonian. As the category of postpositions is traditionally described as including simple members only (Palmeos 1985: 6), the process of grammaticalization

described here amounts to the development of a new (sub)category of function words – the category of complex postpositions. The formal representation of this process is given in Figure 1.

THE SOURCE				THE TARGET			
[*tema-Ø*	*selja-Ø*]	*taga*	>	*tema-Ø*		[*selja-Ø*	*taga*]
s/he-GEN	back-GEN	behind		s/he-GEN		back-GEN	behind
N	N	P		N		P	
simple PP				complex		PP	
'behind his/her back' (lit.)				'after him/her (temp.)', 'behind him/her',			
				'in his/her absence', etc.			

Figure 1. Structure of the simple postpositional phrase (source) and the complex postpositional phrase (the target)

The scope of this study is restricted thematically. This dissertation is chiefly concerned with a small group of postpositional phrases related to body part terms. By 'body part related postpositional phrases' I mean postpositional phrases that consist of a complement noun and a simple postposition. The complement noun refers to a human body part (such as *hand*, *back*, *side*, *foot*). The simple postposition, expressing primarily spatial meanings in these phrases, has itself developed via grammaticalization as well. The function words that serve as postpositions and adverbs in contemporary Estonian have developed from body part terms (e.g. *ear*), parts of objects or animate beings (e.g. *side*), or nouns denoting environmental landmarks (e.g. *field*).

Body part related expressions have received quite a lot of attention in grammaticalization studies. It is well established that body part terms tend to grammaticalize into terms of spatial orientation, and then further into grammatical items that express notions of more abstract domains (e.g. Heine, Kuteva 2002; Heine 1989, Claudi, Hünnemeyer 1991; Svorou 1994; Heine 1997). This tendency is also well attested in Finno-Ugric languages and has been reported in studies within the grammaticalization framework (Ojutkangas 2001; Habicht 2001a; Suutari 2006) as well as in studies that lie outside of this framework (e.g. Palmeos 1985). However, the research thus far has focused on the development of simple grams (*rind* 'breast' > *rinnas* 'abreast' (Habicht 2001b); *käsi* 'hand'[1] > *käes* 'in possession', 'in situation or condition', *külg* > *küljes* 'beside', *pea* 'head' > *peal* 'on', *kõrv* 'ear' > *kõrval* 'next to' (Ojutkangas 2001)). On the other hand, a considerable amount of research on Uralic postpositions has focused on morphologization of postpositions, i.e. the process whereby postpositions turn into case suffixes (Tauli 1966: 112–113). This study concentrates on the development of complex units. Thus, the starting point of the grammati-

---

[1] The body part term *käsi* refers to both hand and arm. For the sake of brevity, it is henceforth translated as 'hand'.

calization process described here is the whole simple postpositional phrase (body part noun + simple (spatial) postposition), and not the postposition or the noun alone (see Figure 1).

Although complex function words have not exactly been in the spotlight of grammaticalization studies, there is a noticeable amount of literature on the (development of) complex adpositions (e.g. Hoffmann 2004; Rostila 2004; Lehmann 1998). However, most of these studies discuss the (development) of complex *pre*positions in English and other Indo-European languages. This account, on the other hand, discusses the development of complex *post*positions and aims to contribute to the ongoing debate by providing evidence from Estonian, a Uralic language.

Grammaticalization is a gradual change in language and, therefore, this process is usually studied diachronically. The diachronic method enables us to observe the dynamics of linguistic phenomena over long periods of time. However, often the data do not allow us to observe the object of study with as much detail as necessary. Thus, although diachronic data have been taken into account as much as possible, this study is primarily a synchronic study. As the phenomenon studied here is an instance of incipient grammaticalization, it enables us to take a closer look at the beginnings of the grammaticalization cycle and witness developments that may be not viable in the long run and may perish along the way of the diachronic grammaticalization path.

The development of complex function words in Estonian is an instance of grammaticalization. However, this process also involves lexicalization, which proves to be a vital stage in this grammaticalization process. Thus, in this study I observe the cooperation of two phenomena which are often considered to work in opposite directions, in the development of complex function words. The presence of both of these phenomena – grammaticalization and lexicalization – enables me to observe the effects of lexicalization on the grammaticalization process.

The present account relies primarily on Habicht and Penjam (2007), who have suggested that the phenomenon under study is part of a larger process – the cyclical development of Estonian function words. The objective is to test the schema of cyclical development of function words on a different set of data, the postpositional phrases related to body parts; to develop a methodology to consistently distinguish complex postpositions and adverbs from the free combination of a simple postposition and its nominal complement; to refine the theoretical rationale of the model, and to establish the development of complex postpositions as an instance of grammaticalization, i.e. establish the criteria of grammaticalization that are at work in this change. Moreover, the model of cyclical development relies upon linguistic experiments (Habicht, Penjam 2007; Jürine 2009; Jürine 2011) and this topic has not been studied with a corpus. In the present study large amounts of real language data are used to describe the synchronic variation as well as the diachronic development of body part related complex postpositions and adverbs in Estonian.

1.2. Theoretical assumptions

This account proceeds from the assumption that the development of functional categories is rooted in grammaticalization. Grammaticalization is defined as a process whereby lexical items become grammatical items, or (less) grammatical items become more grammatical (Heine, Kuteva 2002: 2). Grammaticalization is also a framework that explains the development of functional categories in language (Heine 2003: 587).

The language change that takes place in the grammaticalization process has a cognitive basis (Heine et al. 1991; Heine 1997). That is related to the more general assumption that language reflects the way we perceive the world (inter alia, Janda 2006). One of the main characteristics of cognition is that more abstract concepts are processed through more concrete concepts. Likewise, in language, more abstract notions and categories are expressed with the help of more concrete expressions. For instance, the temporal dimension is often coded in spatial terms. (Heine et al. 1991: 28) However, shifts from one category to another are not random, but rather, they are based on metaphor. As such they represent the most basic categorical shifts that are present in many languages. For instance, body part terms tend to develop into spatial grams (Heine 1989; Svorou 1994), and develop further to express temporal or even more abstract notions (Heine, Kuteva 2002: 183). However, some body part terms may be used to express more abstract notions (*hand* > POSSESSION) without passing through the locative phrase and without expressing spatial reference (Heine 1991: 34).

It is assumed that even though grammaticalization is a diachronic process in essence, it can be studied synchronically. Given that grammaticalization is not a unique process and, therefore, adheres to general principles of language change (cf. Brinton, Traugott 2005: 101), it is also possible to investigate the development of grammatical items in synchronic data. Different usages of a grammatical(izing) item form a synchronic continuum ranging from less grammatical to more grammatical usages. In accordance with general principles of grammaticalization, it can be assumed that the synchronic continuum reflects the diachronic stages of this development. (Heine 1999: 179)

As with language change in general, grammaticalization is considered to be a gradual process occurring slowly over time. It follows that, diachronically, it is not possible to classify the usages into discreet categories (e.g. lexical vs. grammatical usages) in every single case. Likewise, there is no discreet distinction between the lexical and grammatical items synchronically. Lexical and grammatical items are considered to form a continuum, where the lexical items are placed on one end and grammatical items on the other (Brinton, Traugott 2005). Estonian function words are situated in the middle of such a continuum because they exhibit characteristics of both – lexical as well as grammatical items.

Although grammatical and lexical items can be pictured on one continuum, grammaticalization and lexicalization are considered to be neither opposite nor necessarily alternative processes. On the contrary – in this respect, the present account is in accord with Lehmann (2002) and Rostila (2004), who claim that

the development of complex prepositions involves the processes of grammaticalization as well as lexicalization, and Habicht (2001) who also suggests that both of these phenomena can engage in the same process of change.

The study of grammaticalization as, indeed, the study of any aspect of language should rely on data of real language usage. Many authors (e.g. Kemmer, Barlow 2000; Hoffmann 2005) have underlined the importance of the usage-based approach in linguistic research. As grammaticalization is a diachronic process, it is usually studied based on diachronic records. Nevertheless, synchronic studies of grammaticalization, especially of incipient grammaticalization, also have much to gain from corpus linguistic approaches. For instance, large amounts of actual language usage data allow us to observe the token frequency (text frequency) as well as type frequency (pattern frequency) in the observed phenomenon. Frequency is of great importance in many types of linguistic research. As it is considered to play a significant role in the process of grammaticalization (e.g. Krug 1998; 2000; Hopper, Traugott (2003) [1993]; Bybee 2003, 2007, 2010; Hoffmann 2005), it is crucial that one is able to observe its effects when studying the rise of new grammatical items.

1.3. Relevant concepts and terms

In this study, I use the term **'function word'** to refer to adverbs and postpositions. Both of these parts of speech may be structurally simple or complex – there are **simple adverbs** and **simple postpositions** and **complex adverbs** and **complex postpositions**. The term **'simple structure'** is used to refer to usages where the phrases behave as freely combined phrases composed of a body part term and a simple postposition. The term **'complex structure'** is used to refer to usages where the phrases behave as complex units, i.e. complex adverbs or complex postpositions. The term **'hybrid form'** is used to refer to usages that exhibit characteristics of both – the simple structure and the complex structure. Usually, the structure of hybrid forms resembles that of the simple structure, and the semantics of hybrid forms resembles that of complex structures. The term **'in-between case'** is used to refer to cases where it is not possible to decide whether a particular example illustrates a usage of the simple or the complex structure. To avoid over-interpretation of the data, such usages are usually coded as belonging to the simple structure. The term **'body part related phrase'** is used to refer to the phrases under investigation without reference to their lexico-grammatical status.

In the qualitative analysis I have adopted, following many other students of grammaticalization of function words (e.g. Svorou 1994; Ojutkangas 2001), Langacker's (1987; 2008) terms, **trajector** (TR) and **landmark** (LM). According to Langacker (1987: 231) these terms are used in connection with 'relational predications', which also include the Estonian simple function words as well as complex functions words. TR and LM are the linguistic expression of figure and ground organization. TR expresses the more prominent participant of the situation

which is located, evaluated or described, and LM expresses the participant that is characterized as the secondary focus (Langacker 2008: 70), the participant in reference to which the characteristics of TR are specified (Zlatev 2007: 327). In the case of Estonian body part related postpositions, the LM is expressed by the nominal complement of the postposition. In case of simple postpositions, the LM is, therefore, expressed by the body part term (*selg* 'back' in example (1), in case of complex postpositions, the LM is expressed by the nominal complement of the complex postposition (*eestlased* 'Estonians' in example (3)). In case of adverbs, the LM is usually expressed by the (adessive) agent (*autotaltsutaja* Ken Block 'race-car driver Ken Block', example (2). The TR may refer to static or moving persons or objects, and it may also refer to events (Zlatev 2007: 327). In actual language use, the above are not always unequivocally distinguishable, and therefore the notion TR is not used in quantitative analysis.

I use the term **preceding (pro)noun** (PN) to refer to the (pro)noun that precedes the body part related phrase and belongs to the same phrasal structure as the phrase. In cases with the simple structure, the preceding (pro)noun is the modifier of the NP that complements the simple postposition (*tema* 's/he' in the source form, see Figure 1). In cases with complex structure, the preceding pronoun behaves as the complement of the complex postpositional phrase (*tema* 's/he' in the target form, see Figure 1). This term is useful in describing the features that pertain to the simple and complex structures, e.g. the semantic class of the preceding (pro)noun, which help to explain the shift from one structure to the other.

1.4. Research questions

In this study, I seek answers to the following questions:

1. What evidence supports the claim that the freely combined simple postpositional phrases have been reanalyzed as complex structures?

In their model of cyclical development of Estonian function words, Habicht and Penjam (2007) suggest that certain combinations of (pro)nouns and simple postpositions function as complex grammatical units, and should, therefore, be analyzed as complex postpositions.[2] As the category of postpositions has traditionally been considered to include simplex forms only, the proposed development of complex items would suggest the rise of a new sub-category of (complex) grammatical items. Habicht and Penjam claim that the emergence of these complex items is an instance of the process of grammaticalization, which is established by reanalysis of the freely combined simple postpositional phrases to complex postpositions (Habicht, Penjam 2007: 57). Indeed, reanalysis has been frequently associated with grammaticalization as well as lexicalization in

---

[2]   Habicht and Penjam use the term *liitkaassõna* which translates as compound postposition. However, for reasons presented in section 2.3, this study uses the term 'complex postposition' instead.

many accounts (e.g. Hopper, Traugott 2003; Hoffmann 2005). However, reanalysis is a mechanism that is invisible (Langacker 1977: 58), and, therefore, it cannot be directly observed. However, the status of a grammatical item pre-supposes that the item has developed certain features not exhibited by its lexical source form. In the present study, I set out to study synchronically and diachronically whether there is evidence of the 'actualization' (e.g. De Smedt 2012) of the process of reanalysis. In other words, I seek to determine the parameters of grammaticalization involved in the development of complex postpositions and to what extent such parameters are realized in case of each phrase.

2. What is the role of frequency in the development of complex function words? What evidence is there of fixedness and productivity of the phrases under investigation?

The development of complex function words in Estonian has not been examined with a corpus before. Therefore, previous accounts do not include evidence based on frequency. However, the correlation between (high) frequency and grammaticalization has been discussed by many authors (e.g. Bybee 1985, 2003, 2006, 2007, 2010; Krug 1998, 2000; Mair 2004; Hoffmann 2005). High frequency is associated with increasing autonomy (Bybee 2010) as well as productivity (Bybee 1985; 2010; Brinton, Traugott 2005) of the grammaticalizing form. However, it has also been suggested that high frequency is not vital for grammaticalization (Hoffmann 2005) and, moreover, that incipient grammaticalization does not exhibit any significant changes in frequency (Mair 2004). Thus, I set out to study the role of frequency in the development of complex function words. I observe the absolute frequency of the phrases and its correlation with other parameters of grammaticalization, as well as the associational strength between the components of the phrase, which is taken to suggest fixedness of the phrases in question, and associational strength between these phrases and other elements in the immediate sentential context, which is considered to reflect the productivity of the complex items.

3. Is there diachronic evidence of (further) grammaticalization in the development of complex postpositions?

Grammaticalization is a gradual process, i.e. a change that evolves slowly over time, and is, therefore, mostly studied diachronically. Habicht and Penjam (2007) have suggested that the manifestations of the grammaticalization of complex function words are observable in the contemporary language. Accordingly, the phenomenon has thus far been studied only from a synchronic perspective. In the present approach, the parameters of grammaticalization have been defined in such a way as to allow diachronic investigation, enabling me to observe the dynamics of the change and to determine whether usage of the phrases has become more grammatical.

4. Is adverbialization a vital prerequisite of the development of complex postpositions in Estonian?

In contemporary Estonian, it is not uncommon for function words to serve as both adverbs and postpositions. Heine and Kuteva suggest that adverbs are one of the most common sources for adpositions besides nouns and verbs (2007: 83). The model of cyclical development of Estonian function words suggests that when a function word functions as an adverb as well as a postposition, the adverbial stage precedes the postpositional stage. However, this suggestion is based on observations made about simple function words, and its validity with regard to complex function words has not been assessed with a corpus. Moreover, the analysis of contemporary language suggests that in some cases (e.g. *käe all* 'hand+under') the proposed path of grammaticalization is unlikely because of the low frequency and restricted contexts of adverbial usage. Svorou (1994) has also pointed out that frequency should be taken into account when studying the developmental paths of adpositions (Svorou 1994: 105). Therefore, I seek an answer to the question of whether adverbial usage always precedes postpositional usage diachronically.

5. What is the role of lexicalization in the process of grammaticalization of complex postpositions in Estonian?

Habicht and Penjam (2007: 56) suggest that in the cyclical development of function words in Estonian, lexicalization serves as an intermediate stage in the grammaticalization process. They associate lexicalization with the stage of adverbialization and the development of the lexical function (2007: 53). However, as suggested above, the adverbial stage might not be mandatory in the development of complex postpositions. This leads to a more general question of whether lexicalization is a vital step in the development of complex postpositions in Estonian. However, lexicalization has been associated with the development of complex prepositions (e.g. by Lehmann 1998, 2002; Rostila 2004), even without reference to association with the adverbial function. Petré, Davidse, and Van Rompaey (2012) also consider (elements) of lexicalization to be present in the grammaticalization of complex prepositions. Moreover, Brinton and Traugott claim that, even though complex prepositions are adopted into the lexicon, they also exhibit features of grammatical items, and thus, they should be treated as instances of grammaticalization (Brinton, Traugott 2005: 65). Thus, the present account re-considers the role and nature of lexicalization in the development of complex function words.

6. Is the development of Estonian complex function words a language-internal development or a contact-induced change?

The development of Estonian postpositions is generally thought of as the fossilization of case-marked substantive forms (Habicht 2000: 73). As grammaticalization of adpositions from nouns is one of the most common instances of grammaticalization (Lehmann 1991: 501), and because the process follows the general principles of language change, the development of postpositions is usually thought of as a natural, i.e. language-internal change. However, it has been observed that the development of a number of simple postpositions in Estonian is likely the result of contact-induced grammaticalization (Habicht 2000). Considering that there are major similarities in the formation of the complex and simplex forms, and that the beginnings of complex postpositions possibly date back to the same period as the simple postpositions discussed by Habicht (2000), it is possible that German has influenced the development of complex postpositions as well. Thus, the present account investigates whether the development of complex postpositions and complex adverbs shows evidence of contact-induced grammaticalization.

II OVERVIEW OF THE LINGUISTIC PHENOMENON

In this chapter, I give an overview of the linguistic phenomenon – Estonian function words and their developmental paths. The chapter is structured as follows. First, I give a short overview of the Estonian language, touching upon the relevant typological features, such as phrase structure, especially of postpositional phrases. The second section (2.2) focuses on Estonian function words – adverbs and postpositions – and their diachronic development. The third section (2.3) discusses previous treatments of complex postpositions in Estonian. The last section (2.4) gives an overview of the related phenomena in other (both related and unrelated) languages.

2.1. Morphosyntactic structure of Estonian

The Estonian language belongs to the Finnic branch of the Uralic language family (Erelt et al. 2003: 131). According to the census held in 2011, there are 886,859 native speakers of Estonian.[3] Estonian is the official language of Estonia,[4] and accordingly used in all domains, including in education, up to higher education.

The basic word order in Estonian is SVX (subject – verb – object or adverbial) (Erelt et al. 2003: 100). Estonian is primarily a head-final language. A prototypical NP consists of a pre-nominal modifier and a head noun (Erelt et al. 2003: 112) (as in example (4)). In adjective phrases, similarly, the modifiers occur before their heads (Erelt et al. 2003: 116) (see example (5)); modifiers expressing the degree or manner of an adverb in an adverb phrase are also positioned before the head (see example (6)).

(4) *venna-Ø   tool*
     brother-GEN   chair
     'brother's chair'

(5) *küllaltki   huvitav*
     fairly   interesting
     'fairly interesting'

(6) *väga   veidralt*
     very   weirdly
     'very weirdly'

---

[3]   http://www.stat.ee/64629?parent_id=39113 (Accessed 11.01.2016)
[4]   http://www.president.ee/en/republic-of-estonia/the-constitution/index.html
(Accessed 11.01.2016)

Silmilarly, Estonian is predominantly a postpositional language. Although Estonian also makes use of prepositions[5], postpositions are more common (Erelt et al. 2003: 117; 130). This is considered to be a typical feature of Finno-Ugric languages (Palmeos: 1985: 5; Grünthal 2003: 45). As this study is concerned with postpositional phrases, I will henceforth focus on postpositions. Within a postpositional phrase, the majority of postpositions assign the genitive case (Palmeos 1985: 48; Erelt et al. 2003: 117) (as in example (7)). Less commonly, the postposition may take nominals in other cases (elative, nominative, partitive) as well (Erelt et al. 2003: 117–118). This study is concerned only with postpositions that take nominals in genitive case.

(7) *tooli-Ø*   *all*
     chair-GEN   under
     'under the chair'

Typically of Finno-Ugric languages, Estonian is rich in cases. Estonian has 14 cases – three grammatical cases (nominative, genitive, and partitive), and eleven adverbial cases (illative, inessive, elative, allative, adessive, ablative, translative, terminative, essive, abessive, and comitative) (Erelt et al. 2003: 32). It is generally held that Estonian adpositions express similar functions as case suffixes, and that in some instances postpositions and case suffixes are interchangeable (as in (8) and (9)) because they bear exactly the same meaning[6]). However, in most cases, postpositions are not easily replaceable with case suffixes (Erelt et al. 2003: 117). For instance, there is no case suffix equivalent to the postposition *all* 'under' (see example (7)). In general, postpositions are less fixed than the case suffixes. For instance, case suffixes have narrower scope – the scope of postposition may range over several nouns (Palmeos 1985: 4) as in (10), but case suffixes are added to each noun (11).

(8) *laua-Ø*   *peale*
     table-GEN   on.LAT
     'on the table'

(9) *laua-le*
     table-ALL
     'on the table'

(10) *raba-sid,*   *luhta-sid,*   *välja-sid*   *mööda*
     bog-PL.PRT   meadow-PL.PRT   field-PL.PRT   along
     'along the bogs, meadows, and fields'

---

[5] Moreover, some Estonian adpositions are bipositions, i.e. they may be used as both – postpositions and prepositions (Erelt et al. 2003: 118).
[6] Although a more detailed analysis has revealed that there are certain systematic differences in the usage of the adessive case and the postposition *peal* 'on' (for more detail see Klavan 2012).

(11) *raba-de-le,*   *luhta-de-le,*   *välja-de-le*
      bog-PL-ALL   meadow-PL-ALL   field-PL-ALL
      'to the bogs, meadows, and fields'

According to Erelt et al. (2003: 117), there is a subtle line between Estonian case suffixes and postpositions. For example, some cases (such as the terminative, essive, abessive, and comitative) behave more like postpositions in some respects, because in these cases the modifier does not agree with the head (as shown in (12)) as it does with other cases but remains in genitive, as it does in a postpositional phrase (as in (13)) (ibid.).

(12) *ilusa-Ø*   *laua-ni*
      pretty-GEN   table-TER
      'to the pretty table'

(13) *ilusa-Ø*   *laua-Ø*   *peale*
      pretty-GEN   table-GEN   on.LAT
      'on the pretty table'

Thus, Estonian case suffixes and postpositions can be viewed on a single synchronic continuum, ranging from less grammatical (postpositions) to more grammatical (case suffixes) items (cf. Lehmann 1985: 304). Nevertheless, most of the case suffixes cannot be taken to originate from postpositions. Tauli (1966: 12) claims that in the Uralic languages there is a general tendency for case suffixes to be replaced by postpositions instead, especially in the languages with rich case systems. In Estonian, there is evidence of only one case suffix (-*ga*, COMITATIVE) that has developed from a postposition (Habicht 2000: 43–44; Erelt 2009: 19). Tauli notes that the fact that many contemporary adpositions bear the same case suffixes that are used to inflect nouns, makes the development even less probable (Tauli 1966: 12).

2.2. Function words in Estonian: postpositions and adverbs

The Estonian parts-of-speech system can be characterized from three perspectives – semantic, syntactic, and morphological. Based on semantic features, words are divided into autosemantic words (content words) and synsemantic words. Based on syntactic features, words are categorized as autonomous and unautonomous. Based on morphological features, words are either inflected or uninflected (Tauli 1973: 39; Erelt et al. 1995: 18; Erelt 2013: 18). Estonian postpositions are characterized as synsemantic non-autonomous uninflected words. Estonian adverbs, on the other hand, are classified as autosemantic, autonomous, and uninflected words. In the following, I give a brief overview of these two parts of speech.

Estonian postpositions are not considered independent words because they cannot occur by themselves (Habicht 2001: 76). Erelt (2013: 24) has illustrated this claim with the following examples. In (14), the postposition *juures* 'by' is complemented by a compulsory nominal (minu 'I-GEN'); example (15), where the complement is missing, is not grammatical.

(14)   *Ta*   *ela-b*   <u>*minu-Ø*</u>   *juures.*
        s/he   live-3SG   I-GEN   by
        'S/he lives by me.'

(15)   **Ta*   *ela-b*   *juures.*
        s/he   live-3SG   by
        *'S/he lives by.'

Within the framework of cognitive and functional linguistics, function words are usually not considered synsemantic but rather as expressing grammatical meanings. Similarly, in this study I consider postpositions to bear more or less abstract grammatical meanings. Based on their function, Estonian postpositions can be divided into various groups: postpositions of space, time, cause, manner, state, and others, whereby the postpositions may be either mono- or polysemous (Palmeos 1985: 7–8). As this study is concerned with postpositional phrases that consist of body part nouns and spatial postpositions, I will henceforth focus on spatial postpositions. Although Estonian postpositions are considered to be uninflected words, spatial postpositions often come in three form sets, expressing the lative (16), locative (17) and separative (18) (Erelt et al. 1995: 34).

(16)   *käe-Ø*   *alla*
        hand-GEN   under.LAT
        'to under [one's] hand'

(17)   *käe-Ø*   *all*
        hand-GEN   under.LOC
        'under [one's] hand'

(18)   *käe-Ø*   *alt*
        hand-GEN   under.SEP
        'from under [one's] hand'

Synchronically, the lative, locative, and separative forms are considered to be morphologically simple but from the diachronic perspective, the different forms of *all* ('under') include rudiments of locative case endings. According to Wiedemann (1869 [1973]) the lexical source of *all* is the noun *ala* that denotes an area, a region, or a field. Thus, *all*[7] is a postposition which has developed via

---

[7]   I consider *alla*, *all*, *alt* as three forms of the same postposition.

fossilization of a noun in its locative forms, and is considered to be an instance of grammaticalization.

In fact, most of Estonian adpositions have been formed in the process of grammaticalization whereby nouns have fossilized in locative cases (Habicht 2000: 19). In general, Estonian postpositions are rather transparent because in most of the cases the source noun is also still used as a lexical item. Furthermore, in most cases, there has been no alternation (e.g. erosion) to the source form, so that the inflected noun form (as in (19)) and the postposition (as in (20)) co-exist in the contemporary language. It is possible that this withholds the process of grammaticalization to some extent because the presence of the source forms keeps the association with the source form evident. Habicht (2000: 35) has pointed out that the source form of the postposition *kätte* (hand.ILL) 'into ownership, possession' still being used in its lexical sense may hinder the development of *kätte* beyond a secondary postposition.

(19) *Sinu-Ø   inglise-Ø   keel   on   **kõrva-le**   valus   kuul-da.*[8]
      your-GEN   English-GEN language   be.3SG   ear-ALL   painful   hear-INF
      'Your English hurts [my] ears.'

(20) *Istu-Ø   minu-Ø   **kõrvale**!*
      sit-IMP   I-GEN   beside.LAT
      'Sit by me.'

Of course, linking the postposition to the source form does not solely depend on formal similarity. The transparency of a postposition is also dependent on the semantic proximity of the noun and the postposition. The postposition *kõrvale* 'beside' (20) is quite weakly associated with the use of the source form exemplified in (19), and it is difficult to find contexts where the postpositional use would be semantically close to its use with the source form. For instance, the postposition *küljes* 'attached' in example (22) is much more easily associated with its source form *küljes* (side-INE) (21).

(21) *kolm   päeva-Ø   on   ol-nud   kerge   valu   parema-s   **külje-s**,*
      three   day-PRT   be.3SG   be-PST.PTCP   mild   pain   right-INE side-INE
      täpselt   vöökoha-s...
      exactly   waistline-INE[9]
      'For three days there has been mild pain in the right side, exactly in the waist.'

(22) *Konti-de   **küljes**   ol-i   veel   veidi   liha-Ø.*
      bone-PL.GEN   attached.LOC   be-3SG.PST   still   little   meat-PRT
      'There was stil some meat on the bones.'

In addition, association with the source form is also connected with the degree of grammaticalization and the period from which the first attestations as a function word originate. Although there is no proper overview of the degree of grammaticalization of Estonian postpositions, individual studies (Habicht 2000, Ojutkangas 2001, Habicht, Penjam 2007) have indicated that, based on the classification of Lehmann (1985), Estonian has secondary as well as primary postpositions. It must be noted that the development of most Estonian adpositions is not observable in written texts. Habicht has pointed out that by time of the period of Old Written Estonian[10], numerous adpositions (e.g. *all* 'under', *sees* 'in', *peal* 'on') had already reached the status of primary adposition (Habicht 2000: 23). However, the available written materials allow us to observe the development of a group of adpositions (e.g. *pärast* 'because of', *kõrval* 'beside', *asemel* 'in place of', *kombel* 'in the manner of', etc.) that were considered to be in-between cases of nouns and function words in the 16th and 17th century and are considered to function as secondary adpositions in the contemporary language (Habicht 2000: 48–52). There is evidence that some of the postpositions that began their grammaticalization process during this period have evolved due to the influence of German (e.g *asemel* (stead+ADE) 'instead' < *an Statt*, *an Ort*; *heaks* (good+TR) 'on behalf of' < *zu gut*). Thus, the development of some postpositions is caused by translation loans. (Habicht 2000: 51–52).

In addition to the adpositions that were already present by the period of Old Written Estonian, and the adpositions whose development is observable in the diachronic corpora, there are adpositions whose development is observable in present-day Estonian. For instance, the word *korras* 'order' may be interpreted as a noun in adessive case (23) or as a postposition of manner (24) because the form *korras* is dissociating from the nominal paradigm of *kord* 'order'

(23) *Süüdlas-t   karista-t-i   seaduse-ga   ettenähtud   korra-s.*
     offender-PRT   punish-IMPS-PST   law-COM   prescribed   order-INE.[11]
     'The offender was punished as prescribed in the law.'

(24) *Teg-i-n   se-da   käsu-Ø   korras.*
     do-PST-1SG   this-PRT   order-GEN   by.LOC
     'I did it by order.'

Moreover, Sepper (2007; 2006) has described a construction she calls the 'adessive indirectal' which consists of a noun or an NP and another noun in the adessive case which behaves as a postposition (see example (25). Many words may fill that slot (e.g. *sõnu-l* 'word-ADE', *väite-l* 'claim-ADE', *ütluse-l* 'parol-ADE', *hinnangu-l* 'assessment-ADE'). These words refer to someone

---

else's statement, and most of them can only be used in the singular. Such forms can be analyzed as postpositions on the rise. Sepper claims that the frequency of use of such forms increased considerably in journalistic texts in the 1990s (ibid.). Due to their relatively recent appearance, they carry relatively specific meanings (e.g. *hinnangul* 'based on one's assessment'). (Sepper 2007: 312)

(25)   *EKRE-Ø*   **väitel**   *ei*   *vasta-ø*   *Eesti-Ø*   *ja*
EKRE-GEN   according to   NEG   correspond-CONNEG   Estonia-GEN   and
Venemaa-Ø   *piirileping*   *põhiseaduse-le,*   *erakond*   *pöördu-s*
Russia-GEN   boarder agreement   constitution-ALL   party   turn-PST.3SG
hinnangu-Ø   *saamise-ks*   *õiguskantsleri-Ø*   *poole*
assessment-GEN   becoming-TR   chancellor of justice-GEN   to
'According to EKRE the border agreement between Estonia and Russia is not in accordance with the constitution; the party approached the Chancellor of Justice for an opinion.'

Thus, it seems that new adpositions are entering the grammaticalization process quite regularly. The emergence of new adpositions from free combinations of nouns and simple adpositions or case suffixes is considered to be one of the most common cases of grammaticalization in the world's languages (Lehmann 1991: 501). The development of new adpositions in Estonian and the relative difficulty of distinguishing between adpositions and inflected noun forms has been addressed by many authors (Jürine, Habicht 2013; Erelt et al. 2000: 145; Grünthal 2003: 56; Ojutkangas 2001: 47; Jaakkola 1997: 125–129). So it seems that simple grams do not form a strictly closed class of words in Estonian, but rather a semi-open class to which new members are being added all the time.

However, as grammatical uninflected words, postpositions are normally described as parts of speech pertaining to a closed class, i.e. a word class to which, by definition, is not open to easily adding new items (Crystal 2000: 300). Furthermore, as suggested above (in section 2.2), Estonian postpositions are syntactically dependent on their complement and cannot occur alone. Furthermore, their position is fixed – they need to be immediately adjacent to the complement noun. The category of adverbs, on the orhter hand, is considered a lexical, an open category (Erelt et al. 1995: 18). Adverbs specify the verb and are syntactically more independent and positionally free. Because of these reasons, it could be assumed that postpositions exhibit more features of grammatical items and are, hence, further down the grammaticalization path.

However, in Estonian, the two categories stand rather close to each other formally as well as semantically (Jürine, Habicht 2013: 740). For instance, Estonian postpositions and adverbs have exactly the same form; they are distinguished based solely on their syntactic function. Semantically, the two parts-of-speech may not differ as much (see examples (26)–(27)) (c.f. Veismann 2009: 43–45). It is possible that the relative openness of the category of postpositions is due to the semantic closeness as well as formal similarity with adverbs, an open category.

(26) *Kampsun   on   mantli-Ø   **all**.*
    sweater   be.3SG   coat-GEN   under.LOC
               COMPLEMENT   POSTPOSITION
    'The sweater is under the coat.'

(27) *Mu-l   ol-i   mantel peal ja   kampsun   **all**.*
    I-ADE   be-PST.3SG   coat   on   and   sweater   under
                                                        ADVERB

    'I had a coat on and a sweater under [it].'

The formal and semantic 'closeness' of Estonian adverbs and postpositions is conditioned by their mutual developmental path. Estonian has two main ways to form adverbs – syntactic and morphological, the latter involves derivation of adverbs by adding suffixes to mostly adjectival stems (e.g. *hell* 'tender' > *hellasti* 'tenderly'). However, this type belongs to word formation, and is not of concern here. This study deals with adverbs that belong to the syntactic type, i.e. adverbs that have developed via adverbialization. The most common source for such adverbs is nouns (Villup 1969: 33). Similar to the development of postpositions described above, adverbialization affects nouns in locative cases. Adverbialized forms are distinguished from their source forms primarily based on semantic properties. As adverbs, the linguistic items acquire a new, non-literal meaning. Villup (1969: 34) points out that when contrasted with their source form, the meaning of an adverb is always more general. Similarly to postpositions, adverbs often coexist with their source in contemporary Estonian. As can be observed from the examples below, the adverbial use in (28) can be distinguished from the noun form (29) semantically, but not formally. However, the source forms can be inflected, while adverbs usually fossilize in a single form, or in three forms in some cases (spatial adverbs), similarly to postpositions.

(28) *Ta   käi-s   mu-l   eile   **külas**.*
    s/he   go-PST.3SG   I-ADE   yesterday   by.LOC
    'She came by yesterday.'

(29) ***Küla-s**   o-li   kümmekond   talu-Ø.*
    village-INE   be-PST.3SG   about ten   farm-PRT
    'There were about ten farms in the village.'

Semantically, Estonian adverbs are divided into groups expressing similar meanings as postpositions. Thus, there are adverbs of space, time, manner, state, quantity, degree, comparison, cause, purpose, concession, etc. (Veski 1982). Thus, adverbs are not only closely related to nouns as their source forms but also to postpositions, with which they express similar functions. This raises the question of the relationship between the two parts of speech. Indeed, as Habicht (2000) has pointed out, diachronic data suggests that many secondary postpositions were earlier used as adverbs (in periphrastic verbs). Therefore, according to her, adverbialization serves as an intermediate stage in the development of Estonian postpositions.

2.3. Previous treatments of complex adverbs and postpositions in Estonian

Whereas there are numerous studies on the development of simple grams in Estonian as well as in Finnish (Habicht 2001; Sepper 2006; Ojutkangas 2001; Jaakkola 1997), there has not been much research on development of complex grams in Estonian. According to the traditional grammars of Estonian, the category of postpositions consists of simple items only (Palmeos 1985: 6). Adverbs, on the other hand, can be either simple or complex. It has been noted that some complex adverbs 'function' as complex prepositions (Palmeos 1985: 5; Erelt et al. 2003: 91). Thus, it seems that the idea that complex adverbs may be changing their syntactic function similarly to simple adverbs is not entirely new. The idea that Estonian is developing a (sub)category of complex *post*positions was first put forward by Habicht and Penjam (2007)[12]. They have proposed a schema that depicts the evolution of function words as a cyclical process (see Figure 2).
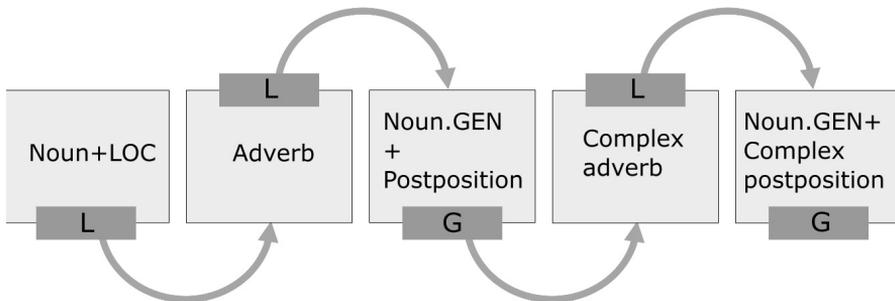
Figure 2. Cyclical evolution of function words in Estonian (adapted from Habicht, Penjam 2007: 57)

---

[12] Habicht and Penjam (2007) use the terms 'compound adverb' and 'compound adposition'. Due to a slight difference in perspective on this process, the terms 'complex adposition' and 'complex adverb' are preferred in this account. According to Habicht and Penjam (2007), the development of complex function words is manifested by the tendency to write adpositional phrases as a single word. However, they claim that spelling (alone) does not constitute further grammaticalization but is rather a hallmark of 'a certain type of lexical stage' (Habicht, Penjam 2007: 56–57). Jürine (2011) who has investigated the link between spelling and grammaticalization, reports an association between non-literal interpretation of Estonian adpositional phrases and spelling. However, in the present account, the factor of spelling is not investigated. As it is not a primary criterion of linguistic analysis (cf. Lieber and Štekauer 2009: 7–8), it is a problematic indicator in many respects and, therefore, not comparable to other criteria observed here. Thus, even though some of the studied phrases are occasionally written as a single word (in some cases 'incorrectly', in some cases according to the standard), this facor is ignored here.

According to Habicht and Penjam, the process starts out from a lexical item, i.e. a noun in a locative case which bears a lexical (L) and a grammatical (G) function; the former is expressed by the lexeme and the latter by the case suffix (stage 1 in Figure 2). By stage 2, the grammaticalizing item functions as a simple adverb and bears a lexical function (L). This is characterized as the intermediate stage of lexicalization in the grammaticalization process (Habicht, Penjam 2007: 56). By stage 3, the postposition has combined with a nominal, functions as a postposition and, therefore, carries a grammatical function (G). Note that in stage 3, we are dealing with a fully compositional syntactic phrase. By stage 4, the formerly freely composed postpositional phrase has developed an independent meaning (or several meanings) and hence it has lexicalized and become a holistic unit. As such it functions as a complex adverb and bears, once again, a lexical function (L). By stage 5, the syntactic function of the complex item has changed, so that in combination with a nominal, it forms a complex postpositional phrase where it, once again, carries a grammatical function. In examples (30)–(34), this model is illustrated with one of the expressions investigated here – *käe kõrval* (hand+by), the development of which started out with the adessive form of the body part term *kõrv* 'ear' (30). The inflected noun *kõrval* passed through the stages of an independent adverb (31), and a simple postposition (32), and is currently also used as a complex adverb (33) as well as a complex postposition (34). Note that all of these usages are present in contemporary Estonian.

(30) *Sääsk   maandu-s   otse   lapse-Ø   **kõrva-l**.* (constructed example)
       mosquito   land-PST.3SG   right   child-GEN   ear-ADE
       'The mosquito landed right on the child's ear.'

(31) *Algul   istu-si-n   lapse-l   **kõrval**   ja   teg-i-me   koos*
       at first sit-PST-1SG   child-ADE   beside   and   do-PST-1PL   together
       kõik   kodutöö-d.[13]
       all   homework-PL
       'At first I sat next to my child and we did the homework together.'

(32) *Eile   rohi-si-n   tuul-t   nauti-des   päevaliilia-te*
       yesterday weed-PST-1SG   wind-PRT   enjoy-GER   lily-PL.GEN
       peenar-t,   kui   jä-i-n   korraga   vaata-ma,   et
       flower bed-PRT   when   stay-PST-1SG   suddenly   look-SUP   that
       *mis   konksu-d   need   just   mu-Ø   **käe-Ø**   kõrval   püsti*
       what   hook-PL   these   right   I-GEN   hand-GEN   beside   up
       on,   küüne-d   küljes?[14]
       be.3SG   claw-PL   attached.LOC
       'Yesterday I was weeding the lilies, and noticed that there were some hooks with nails that were sticking up just next to my hand.'

---

[13]   http://www.sinamina.ee/ee/noustamine/e-noustamine/postitused/?fid=14&tid=595&show_all=1 (Accessed 11.01.2016)
[14]   http://koerakodu.wordpress.com/2010/08/05/05-08-2010/ (Accessed 11.01.2016)

(33) *Mõne-Ø*   *tunni-Ø*   *pärast* *tule-b*   *mees* *jälle* *vastu,*
few-GEN   hour-GEN   after   come-3SG man   again opposite
ahv   ***käekõrval.***[15]
monkey   hand.beside.LOC
Lit. In a few hours the man appears again with the monkey beside his hand
'In a few hours the man appears again with the monkey by [his] hand.'

(34) ***Etnoloogi-Ø***   ***käekõrval***   *India-sse* *ja* *Hiina-Ø.*[16]
Ethnologue-GEN hand.beside.LOC India-ILL and China-ILL
Lit. To India and China beside the hand of an ethnologist
'To India and China with an ethnologist.'

The schema proposed by Habicht and Penjam captures the structural changes that the function word is subjected to throughout the whole cycle. In this study, the discussion is limited to stages 3 to 5, which describe the evolution of the complex adverbs and complex postpositions. While my account relies heavily on the above model, the focus of this study is somewhat different than that of Habicht and Penjam (2007). Thus, I will accept the model with a few adjustments see section 2.5).

2.4. Related phenomena

2.4.1. Similar phenomena in Uralic languages

There are no known instances of complex postpositions in other languages related to Estonian. However, there is another, similar phenomenon – suffixation of postpositions – that is reported to have occurred in a number of Uralic languages (Tauli 1966: 112). Tauli lists a number of instances reported from various Uralic languages where postpositions have become case endings. Tauli claims this is caused by the agglutinative tendency of the Uralic languages (ibid.). The most widespread instance is the development of comitative case in the Finnic languages. This phenomenon has also taken place in Estonian, where the postposition *kaasa* 'with, together' has developed into a comitative case marker (*-ga*). According to Tauli (1996: 112–113), the same development has been reported in many Finnic languages, such as Livonian, Votic and Vepsian.

In other related langues, other postpositions have become a comitative suffix, such as *kuim*, *guim* 'companion' in Sami, *taʾịl* 'full' in Mansi and in Khanty. Besides the comitative, in the north-eastern group of Finnic languages, postpositions meaning 'near, by' have agglutinated to the genitive form of the noun and become case suffixes, which are subject to vowel harmony (Tauli 1966: 114–115). In Votian, Ingrian, Ludian and Vepsian the terminative suffix

---

developed when a function word meaning 'up to, till' agglutinated to the illative form of the noun. In Ludian and Vepsian, a postposition *päin* 'towards; from' has been reported to have agglutinated to noun in illative, ablative, and allative forms. In Vepsian, the postpositional forms *алӡ* 'under, beneath' and *рäle* 'on' are used in parallel with the agglutinated reduced forms -*ал*, *ā* and -*рäӡ, рā̄*. The suffixation of postpositions is most extensive in Hungarian, in which the agglutination process is supported by phonetic factors that contribute to the assimilation of the postpositions (Tauli 1966: 116–117). Spencer (2008) is of the opinion that the so-called Hungarian cases should not be considered cases at all, because their function is often highly specific and they do not bear any real grammatical functions. According to him, it would be more appropriate to treat them as 'fused postpositions' or 'regular portmanteaux'. Krista König (2011), who has described the development of Hungarian case suffixes from the perspective of grammaticalization, has characterized these forms as being on their way from an adposition to affix.

The processes described above seem to be restricted in the sense that these processes only result in certain cases (the comitative), and certain nouns/post-positions (e.g. *saadik* 'since', *kanssa* 'people with, together', and its cognates in related languages). As such, these instances are of a different nature than the phenomenon studied here. In Estonian, at least at this point, this change does not seem to be limited to any single postposition or single function.[17] The motivation for the development of complex function words in Estonian is lexicalization, i.e. the phrase consisting of a nominal and a postposition acquires a new, independent meaning. The phenomenon described above stands for the development of case suffixes out of simple grams. This process can be characterized as MORPHOLOGIZATION (cf. Kabak 2006; Haspelmath 2011), which follows the cline CONTENT WORD > GRAMMATICAL WORD > CLITIC > INFLECTIONAL AFFIX (Hopper, Traugott 1993: 7). However, in this case, the result of the phenomenon studied here is a complex grammatical word that may function as an adverb as well as a postposition but by no means a case suffix (at least for now).

However, there are certain similarities between the two phenomena – both of these are possible due to the agglutinating tendency present in Uralic languages. The suffixation of the postpositions is, of course, not confined to Uralic languages alone. For instance, Kabak (2006) has described a similar development in Turkish postpositions. Noonan (2008: 134) has described the origin of the case markers in Bodic languages where the NPs have given rise to

---

[17] Though there are a few Estonian postpositions (*kaupa*, *viisi* (both: 'by'), *väel* 'in' 'wearing only X') that agglutinate to the genitive stem of the nominal producing forms expressing the same function (as in *kaheksakaupa* 'by eight', nädalakaupa 'by the week'; *tükiviisi* 'by the piece' *tosinaviisi* 'by the dozen'; *särgiväel*, 'in one's shirt', pluusiväel 'in one's blouse', mantliväel 'in one's coat'). However, despite the postposition 'glued to' the nominal, they cannot be characterised as suffixes on the basis of any other feature for there is no erosion, the forms are semantically transparent, and even interpretable as morphologically complex.

postpositional phrases as the nominal head in the locative case have grammaticalized into a postposition, and from there on into a case clitic. However, as this study regards a cyclical development where new postpositions arise from already existent simple postpositions, I now turn to such phenomena in other languages.

2.4.2. Similar phenomena in non-related languages

The development of complex function words, which includes the development of complex adpositions (in this case, postpositions), is not, of course, confined to Estonian language. On the contrary, quite a number of studies report similar processes taking place in other, mostly Indo-European languages. Most of the studies known to me focus on complex prepositions. However, some studies also discuss the development of complex postpositions.

Although there is some literature on the topic of (grammaticalization of) complex postpositions, there are not many studies that focus on the determining and development of complex postpositions (with exceptions of Rhee (2004) and Choi-Jonin (2008)). For instance, Rhee (2004) lists a number of criteria to distinguish primary and secondary postpositions in Korean, e.g. morphological complexity, morphosyntactic and phonological changes, orthographic changes, and functional and semantic specialization. These criteria are re-evaluated in Choi-Jonin (2008). However, most of the studies on complex postpositions tend to be either descriptions of single instances of grammaticalization or treatments of complex postpositions with respect to other, often practical, research goals. For instance, Ahn (2010) describes the grammaticalization of the causal complex postposition *tekpwuney* 'thanks to' in Korean. Morphologically complex forms that behave as postpositions are also mentioned, for instance, in Svorou (1994), Heine, Kuteva (2002). Because complex postpositions, like complex prepositions or any multi-word units pose a problem for natural language processing, there are a number of studies that approach complex postpositions from this perspective. For instance, de Ilarraza, Gojenola, Oronoz (2008) discuss Basque complex postpositions from the point of view of developing a system that checks for the correct use of complex postpositions, and Arriola (2012) is concerned with the syntactic disambiguation of complex postpositions in Basque. However, Basque complex postpositions are not structurally similar to Estonian complex postpositions, as the former exhibit a pattern suffix$_1$+lemma+ suffix$_2$, which behaves like an 'appended' complex case suffix (de Ilarraza, Gojenola, Oronoz (2008: 31)). Moreover, Arriola (2012: 5) notes that the distinction case/postposition is disputed in Basque. Similarly, one of the few complex adpositions that behaves as a complex postposition in German – *zu folge* 'according to' (Lehmann 1991: 503) – is structurally different from Estonian complex postpositions because its source form includes a preposition. Because of the differences in research goals and the structure of the items studied, it is rather difficult to contrast the development of Estonian complex post-

positions to those described above. However, there is a fairly substantial body of literature on the development of complex *pre*positions. In the following, I give a brief overview of these studies as well as critical approaches to the development of complex prepositions.

The debate has probably been most extensive on the English complex prepositions. This topic has been addressed, for instance, in Hoffmann (2005), McMichael (2006), Waters (2009) and more recently in Smith (2013) and Petré, Davidse, and Van Rompaey (2012). Hoffmann (2005), who represents a corpus linguistic approach to the grammaticalization of complex prepositions in English, argues for the existence of complex prepositions as complex grammatical items and uses large corpora to investigate the development of complex prepositions diachronically as well as synchronically. Smith (2013) investigates the variability within the complex prepositions (NPN-constructions) focusing on the presence of the determiner within the utterance (e.g. *in (the) light of*, *at (the) risk of*), and its impact on the fixedness, decategorization, semantics and degree of grammaticalization of the construction. The question of categorization has also been raised in connection with complex prepositions by Petré, Davidse, and Van Rompaey (2012) who analyze constructions related to the 'way-nouns', e.g. *en route to*. Andrew McMichael (2006) has studied a different type of complex prepositions, namely what he calls *compound* prepositions in English. He states that the compound prepositions form a large group which includes such elements as e.g. *behind*, *before*, *between*, *about*, *across*, *after* that have grammaticalized from the phrase structure Preposition + NP. Cathleen Waters (2009), who represents a more formal approach to the topic, describes the English prepositions of various structures as representing three stages of the same grammaticalization process that occurs in a cyclical manner. In her account, the prepositions with the structure PNP are currently in the first stage: the spatial item and the noun have combined into an innovative prepositional element which carries a meaning which is not the sum of its components (such as *in front* (*of*)). In the second stage, the grammaticalizing element appears orthographically as a single word (*inside the house*). In the third stage, the preposition is reanalyzed as a single item; (*behind*). In addition to the above mentioned studies, English complex prepositions have been addressed in many other accounts that address various topics related to grammaticalization (e.g. Becker, Bybee 2009; Schwenter and Traugott (1995); Heine, Kuteva (2007); Brinton, Traugott (2005), Svorou (1994)).

The development of complex grams has been discussed in other Indo-European languages, such as German and French. The German complex prepositions have been discussed in German(ist) literature (e.g. Lehmann, Stolz (1992) but also in English (e.g. by Lehmann (1998), Rostila (2004) and (2006); Trawiński (2003). Rostila (2004) discusses the development of the complex preposition *(in) Richtung*, which he describes as being an instance of lexicalization as well as grammaticalization. Trawiński (2003) who represents a formalist approach to complex prepositions, analyzes German complex prepositions with the pattern PNP (e.g. *auf Grund von* 'by virtue of'). Lehmann

(1998), who has discussed the abstract complex prepositional phrases in German, distinguishes the following stages in the development of complex prepositions from verbal nouns: abstract postpositional phrase (*nach Lösung* 'after solution'), prepositional locution *in Ermangelung* 'for want', and complex preposition (*im Laufe* 'during', *mithilfe* 'by means of'). A similar distinction is also made in French. Silvia Adler (2008), who has discussed the criteria to distinguish three different prepositional expressions in French – prepositional phrases, prepositional locutions, and compound prepositions – has pointed out that the latter two are lexicalized uses, while the former is not. She distinguishes the prepositional locutions (e.g. *à l'instar de* 'like') and compound prepositions (e.g. *au mépris de* 'despite') based on their tolerance of syntactic modification. Moreover, there are studies on complex prepositions in Brazilian Portugese (Sheperd 2014) as well as in Spanish (Lehmann 2002).

Moreover, there are studies that describe a similar phenomenon in yet other, Germanic languages, such as Dutch and Swedish. Moirón and Bouma (2003) have discussed the so-called 'collocational prepositional phrases' in Dutch (e.g. *ten opzichte van* 'with respect to', *in plaats van* 'instead of', *in antwoord op* 'in reply to'). The collocational prepositional phrases are considered to be semantically non-compositional and syntactically rigid or idiosyncratic, which is taken to distinguish them from regularly built prepositional phrases (Moirón, Bouma 2003: 153; Bouma, Villada 2002: 23). The authors implemented various statistical tests (mutual information, chi-square test, log-likelihood) to determine the candidates for possible fixed expressions, which they then presented to human judges to be evaluated with respect to fulfilling relevant linguistic properties. However, the authors report great differences between judges and conclude that identifying potential collocational prepositional phrases is a difficult task (Moirón, Bouma 2003: 157–158). Sigurd (1993), who takes the perspective of formal grammar and automatic translation, suggests that there are a number of utterances (e.g. *i mitten av* 'in middle of', *på grund av* 'because of', *i fråga om* 'as to') in Swedish that should be considered 'multi-word prepositions' and included in the lexicon as a separate entry. He establishes a number of linguistic criteria to determine the status of a multi-word preposition and presents a list of complex prepositions in Swedish, based on meeting one or more of the criteria. The prepositions listed belong to various semantic classes (e.g. location, tempus, cause, etc.).

Although the development of complex adpositions seems to be quite a widespread phenomenon, and the class of complex prepositions seems to be quite substantial and 'ever expanding' (Petré, Davidse, and Van Rompaey (2012)), their existence has been called into question by some. Seppänen et al. (1994) who have been, perhaps, the harshest critics of English complex prepositions, have criticized the treatment of these elements solely based on semantics and suggest that the status of PNP strings should be determined by constituency tests. They implement the tests of fronting, coordination, ellipsis, and interpolation and conclude that the English complex prepositions do not exist. Although almost none of the tests implemented by Seppänen et al. (1994) can

be applied in the present study, it is clear that Estonian complex postpositions also allow several interpretations, including that of freely combined phrases. However, in this account, this is not taken to suggest that Estonian complex postpositions do not exist. Grammaticalization is taken to be a gradual process, and the ability of the grammaticalizing item to allow several syntactic interpretations is natural, as grammaticalization often involves layering (Hopper, Traugott 2003 [1993]). Moreover, Hoffmann (2005) argues that the criteria on which Seppänen et al. (1994) base their argumentation is marginal in his data and that these instances do not disprove the existence of complex prepositions. Following Hoffmann (2005), it is assumed in this study that while language change is a gradual process which is dependent on many criteria, a viable account should rely on as many criteria as possible, and in order to avoid over-estimation of the role of marginal usages, also account for frequency.

Although the above-mentioned studies have offered me inspiration and theoretical support for studying the development of complex postpositions, they cannot serve as direct models when working out my methodology for determining the complex postposition in Estonian because of structural differences between the languages. Thus, my account of Estonian complex postpositions relies heavily on Habicht and Penjam (2007), who have proposed a model for Estonian function words. Wherever possible, the criteria suggested in the studies above have been the basis to elaborate the methodology used on Estonian complex postpositions. Nevertheless, most of the criteria or syntactic properties suggested for distinguishing complex prepositions cannot be implemented in the case of complex postpositions, with the exception of the principle of restricted modification, which has been mentioned in several studies (e.g. Moirón, Bouma 2003: 157; Sigurd 1993: 202) (see section 2.5.3.3). Moreover, although the distinction between the prepositional locution and complex preposition may be relevant in the case of German and French, as it helps to build bridges between the simple structure and complex structure, this distinction is not used in the present study because the Estonian complex postpositions are still in their very early stages. Because of this, a simpler, dichotomous division is used. The dichotomy is, of course, an oversimplified model of gradual change where less grammatical forms (pertaining to simple structure) and more grammatical forms (pertaining to complex structure) form a continuum.

2.5. The present account on the development of complex adverbs and complex postpositions

As the development of complex adverbs and postpositions is, at this point, still in its early stages, the development of complex units is synchronically characterized by layering (c.f. Hopper 1991: 21–22, Hopper, Traugott 2003 [1993]: 123–124). In contemporary language, the simple and the complex structures – the source form and the current result – co-exist. Thus, the phrases studied here occur in contexts were they may be analyzed either as freely

combined phrases, i.e. simple postpositional phrases, or as complex units – adverbs or postpositions. As this type of divergence is typical for grammaticalization (Hopper, Traugott (2003) [1993]: 117), it is also attested among some of the simple postpositions (c.f. examples (19)–(20), (21)–(22) and (23)–(24) in section 2.2). It is one of the aims of this study to identify the criteria to properly distinguish the simple structures from complex structures.

Based on previous research, the general features of grammaticalization and lexicalization, and observations of the object of study in the contemporary language, a number of features have been identified in the use of the body part related postpositional phrases. I argue that these features suggest lexicalization and grammaticalization, and are, thus, appropriate criteria to distinguish between the simple and the complex structure. These features are **unit interpretation, extension beyond human reference,** and **non-agreement of the body part term and the preceding (pro)noun**. The present study also accounts for **frequency**, which has often been associated with grammaticalization. In this account, absolute as well as pattern frequency is used to observe **fixedness** and **productivity** of the phrases studied. Moreover, pattern frequencies of the phrases are used to observe how widely spread the above-mentioned features are among the phrases.

This chapter will explain how these features are realized in the development of complex function words and how they indicate lexicalization, reanalysis, actualization, and grammaticalization. The chapter is structured as follows. First of all, I discuss the role of lexicalization and explain the relationship of lexicalization and grammaticalization in the grammaticalization process studied here. Secondly, I will demonstrate why the development of the complex postpositions is considered to be a process of grammaticalization by showing that the freely combined phrases have been reanalyzed as complex postpositions, and that the features listed above constitute an actualization of this process. Then, the role of frequency and its implications will be discussed.

2.5.1. Lexicalization

One of the key criteria for identification of the complex items is unit interpretation. The source of the complex item is a simple postpositional phrase, i.e. a freely combined syntactic construction that consists of a postpositional head and its nominal complement. As a simple structure, the phrase expresses a compositional meaning. As the phrase develops into a complex unit that is stored in the 'mental' lexicon, it comes to express meaning(s) that are not directly derivable from the components of the phrase, and that are carried by the whole utterance. As such, the phenomenon is in line with the definition of lexicalization proposed by Brinton and Traugott:

Lexicalization is the change whereby in certain linguistic contexts speakers use a syntactic construction or word formation as a new contentful form with formal and semantic properties that are not completely derivable or predictable from the constituents of the construction or the word formation pattern. Over time there might be loss of internal constituency and the item may become more lexical. (Brinton, Traugott 2005: 96)

According to Brinton and Traugott (2005: 96–97) lexicalization also includes idiomaticization and fusion. Both of these phenomena are attested in the development of complex function words in Estonian. For example, consider the utterance *mehe+moodi* (man+like) which, in contemporary language, functions as both – a simple postpositional phrase (as in example (35)) and a complex adverb (as in example (36)). In example (35), the utterance is considered to be an instance of the simple structure because the first component *mehe* (man.GEN) refers to a concrete referent ('my husband'). However, in example (36) *mehemoodi* (man+like) is idiomaticized, and the constituents of the phrase lose their compositionality. Hence, the utterance is analyzed holistically, and it functions as an adverb of manner expressing the meaning 'a lot'. In this case, the morphological boundaries have become vague and the utterance has become more fused. It should be noted that in example (36) *mehemoodi* is still transparent. However, following Brinton and Traugott (2005) who state that a lexicalized form can also be complex, I consider the shift from phrasal level to lexeme to indicate fusion (cf. *out-of-hand* in Brinton, Traugott 2005: 97).

(35) *Ja   lapse-Ø   välimuse-Ø   põhjal   nüüd   küll   minge-i-d*
    and   kid-GEN   appearance-GEN   based   now   even   some-PL-PRT
    järeldus-i   teh-a   ei   anna-Ø,   minu-Ø   laps
    conclusion-PL.PRT   make-INF   NEG   give-CONNEG   I-GEN   kid
    ei   ole-Ø   mitte   ühe-st-ki   otsa-st   minu-Ø
    NEG   be-CONNEG   not   one-ELA-CL   end-ELA   I-GEN
    ***mehe-Ø   moodi**, aga   tema-Ø   oma   kahtluse-ta.*[18]
    man-GEN   alike   but   s/he-GEN   own   doubt-ABE
    Lit. You should not draw any conclusions based on the appearance; my child is not like my man in any way but is his without any doubt.
    'You should not draw any conclusions based on the appearance; my child does not resemble my husband in any way but is his without any doubt'

(36) *Kiirtoit,   mis   **mehemoodi**   toida-b.*[19]
    fast food   that   man.like   nourish-3SG
    Lit. Fast food that nourishes like a man.
    'Fast food that nourishes well.'

---

[18]   http://naistekas.delfi.ee/foorum/read.php?9,1421045,page=2 (Accessed 11.01.2016)
[19]   http://epl.delfi.ee/news/kultuur/kiirtoit-mis-mehemoodi-toidab.d?id=51298232 (Accessed 11.01.2016)

It should be noted that usages as illustrated in example (36) are not instances of regular word formation, which are usually not considered to be instances of lexicalization (c.f. Brinton, Traugott 2005). According to Kasik (2013) the process of development of compound adverbs in Estonian is not considered to be productive. Moreover, Erelt et al. state (1995: 597) that the formation of compound adverbs does not involve composition proper because composition as a means of word formation usually presupposes that the second component of the compound expresses the basic meaning of the compound and determines the part of speech (as in example (37)). However, the development of complex function words is the result of two words 'melting' together as they develop a new meaning that is not the sum of the components (as in example (38)). Thus, the development of Estonian complex function words is not described as an instance of word formation, but as a process whereby a new lexical form arises gradually, hence, lexicalization.

(37)　*magus*　+　*hapu*　>　*magushapu*
　　　 sweet　　　 sour　　　 sweet-and-sour

(38)　*vahe*　+　　　　　　　　　　 *peal*　>　*vahepeal*
　　　 gap (space separating something) on　　　 between, meanwhile, sometimes
　　　 noun　　　　　　　　　　 postposition compound adverb

Habicht and Penjam (2007) also regard lexicalization as a prominent process in the development of compound adverbs and postpositions. For them, lexicalization stands for the development of a new item that carries a lexical function (i.e. an adverb). On the other hand, it also stands for the development of a new uninflected word (2007: 53). This means that in the sense of Brinton and Traugott (2005), they consider it to be lexicalization in the narrow sense (i.e. the development of a lexical item) as well as the broader sense (adoption into the inventory). As in this study, the former process is called into question (see section 1.4. and 4.8.2), I only adopt the broader definition of lexicalization, i.e. adoption into the inventory. Nevertheless, in the present account, it is maintained that lexicalization is required here as a prior condition for grammaticalization. If the phrases were not adopted into the inventory, the non-literal readings of *mehe moodi* (man+like) (in example (36)) could be accounted for by the polysemy of the components. However, the components of the phrase do not carry such meaning as single units (for instance *mees* 'man' does not express the meaning 'a lot'). Thus, the meaning of the phrase is not directly derivable from its components. Therefore, it is argued that lexicalization is part of the grammaticalization process of complex function words in Estonian and operates in connection with both – the development of a lexical item (complex adverb) and the development of a new grammatical item (complex postposition) (for a similar approach see Lehmann (2002), Rostila (2004); (2006).

　　Considering the development of the complex unit as an instance of lexicalization without consideration for the part of speech may raise a question of the

placement of complex adverbs and complex postpositions relative to each other on the lexis-grammar cline. It was pointed out when describing simple function words (section 2.2) that the categories of adverbs and postpositions are rather close in Estonian, and that in some cases they may express meanings that are rather similar. Jürine and Habicht (2013: 740) point out that this especially applies to complex items because, in general, they are more recent, and the adverbial (39) and the postpositional (40) uses have not yet diverged to any great extent. However, it is expected that the complex postpositions will drift away from complex adverbial uses in time, as they undergo further developments associated with grammaticalization (e.g. extension to new contexts, increase in frequency and productivity, etc.). Some of these features of grammaticalization are already observable among postpositional uses (see section 2.5.3ff).

(39) *Ja   see   ei   tule-Ø   vaid   selle-st,*
      and   this   not   come-CONNEG   only   this-ELA
      *et   need   ei   satu-Ø   tihti   **käe-Ø**   **alla** /…/*
      that   these   not   appear-CONNEG   often   hand-GEN   under.LAT
      Lit. And this is not only because these [books] do not appear often under hand.
      'And this is not only because one does not come into new books often.'

(40) */---/ siis   nagu   just   selle-ks   puhu-ks   hüppa-s   **mu-Ø***
      then   like   just   this-TR   occasion-TR   jump-PST.3SG   I-GEN
      ***käe-Ø**   **alla**   ümbrik,   mille-Ø   peal   ol-i*
      hand-GEN   under.LAT   envelope   what-GEN   on   be-PST.3SG
      suurelt   LEPING   kirjas.
      large   contract   written down
      Lit. Then, as if for the special occasion, an envelope with the word CONTRACT written on it, appeared under my hand.
      'Then as if for the special occasion, I found an envelope with the word CONTRACT written on it'

As adverbs and postpositions, the studied phrases express more abstract meanings than as freely combined phrases. As such it is indicative of desemanticization, which is considered to be one of the key parameters of grammaticalization (Heine, Kuteva 2002, 2007). The present instance of semantic change is considered to indicate demanticization because when used as a holistic unit the components, especially the body part term as the more contentful component, is losing its referential capacity. For instance, *mehemoodi* in example (36) expresses lexically less contentful meaning than *mehe moodi* (35). The same applies to *vahepeal* in example (38) and *käe all* (hand+under) in examples (39) and (40). However, desemanticization is not considered to be a mere loss in semantic substance but rather a rise of new aspects of meaning (Heine, Kuteva 2006: 60). Accordingly, the usages of postpositional phrases that have gone under desemanticization exemplified above are considered to have gained abstract and, therefore, more grammatical meanings.

2.5.2. The relationship of lexicalization and grammaticalization

Lexicalization and grammaticalization are considered to co-occur in a single pathway of language change by several authors. For instance, Lehmann (2002: 1) suggests that in the case of the development of function words, such as complex prepositions, grammaticalization is always accompanied by lexicalization. In his approach, lexicalization, which stands for holistic interpretation, is needed for the grammaticalization of complex items to occur (Lehmann 2002: 15–16). Similarly, Rostila (2004: 1) suggests that lexicalization is involved as a 'preparatory factor' in the grammaticalization process of the German complex preposition (*in*) *Richtung*. According to Rostila (2004: 3) lexicalization facilitates the grammaticalization process in that it helps to lose the referential capacity of the components of the complex item.[20] Losing the referential capacity of the components leads to demotivation of the complex item, which is considered to be characteristic of lexicalization as well as grammaticalization (cf. Brinton, Traugott 2005: 105). In fact, many other accounts of the development of complex function words make reference to holistic interpretation of complex adpositions. For instance, Hoffmann (2005) uses 'unit interpretation' as one of the key criteria to argue for the existence of complex prepositions in English (2005: 57). Moirón and Bouma (2003: 153) claim that the semantics of the 'collocational prepositions' in Dutch is non-compositional; Bouma and Villada (2002: 6) state that the idiosyncratic properties suggests that they must be 'at least to some extent' lexicalized; Adler (2008: 20) claims that the French compound prepositions and prepositional locutions are lexicalized. Moreover, the co-occurrence of lexicalization and grammaticalization is not confined to the development of complex adpositions – Habicht has suggested that these two processes work hand in hand in the development of verbs into discourse particles and verbs (Habicht 2001: 75).

However, Brinton and Traugott (2005: 65) and Traugott (2003: 636) treat the development of such complex items as grammaticalization (only), because they exhibit features of grammatical items (e.g. syntactic reanalysis and decategorialization). Although they consider lexicalization and grammaticalization to involve a number of mutual parameters (e.g. gradualness, unidirectionality, demotivation), they state that lexicalizations results in items that belong to an open class and grammaticalization results in closed-class items. (Brinton, Traugott 2005: 100).

In accordance with Brinton and Traugott (2005: 105–109), the present account consideres lexicalization and grammaticalization to be gradual and unidirectional processes, which are accompanied by demotivation of meaning. In course of lexicalization, demotivation is manifested in increase of idiosyncracy, lexical content, and specificy. In case of grammaticalization, on the

---

[20] In Rostila (2006), his account on the same topic is further developed and represented in the Construction Grammar framework, where the term lexicalization has been substituted with 'storage' but its role in the process remains the same.

other hand, demotivation of forms leads to more abstract meaning. Claiming that the grammaticalization of complex postpositions involves both of these processes seems to be incompatible with the unidirectionality of grammaticalization. However, as lexicalization and grammaticalization are not considered to operate as opposite phenomena or mirror images (Brinton, Traugott 2005: 62–63 Lehmann 2002: 1), their co-occurrence does not challenge the unidirectionality claim.

This study is concerned with the development of complex adverbs and complex postpositions that are both in the middle of the lexicon-grammar cline, placed rather close to each other, and thus, also on the border of open and closed categories. The development of such units is not considered a change that affects a single item towards a more lexical or grammatical use but a change, whereby a syntactic phrase gradually becomes a single holistic unit. As such, it can be used as an adverb or a postposition and it may then, develop further characteristics consistent with grammaticalization alone (e.g. decategorization, bleaching, productivity, frequency, and typological generality) (Brinton, Traugott 2005: 105–109).

2.5.3. Reanalysis

Another mechanism that needs to be discussed in connection with the development of complex grams is reanalysis. However, reanalysis is a somewhat controversial notion, at least with respect to its connection with grammaticalization. Some authors consider grammaticalization and reanalysis to be entirely different phenomena, as they have found theoretical difficulties in the compatibility of these notions, especially regarding the nature of the two processes. Namely, grammaticalization is thought of as a gradual change which occurs over time, whereas reanalysis is traditionally thought of as an abrupt change (Haspelmath 1998; De Smet 2012). I will return to the problem of abruptness of change below, in section 2.5.3.3.

Some authors have excluded reanalysis from the list of mechanisms relevant to grammaticalization. The reason for this is that the notion of reanalysis has been understood in a number of different ways by various authors. For instance, Heine and Kuteva (2002: 5) claim that whether grammaticalization involves reanalysis or not, is a theory-dependent issue. Although Heine and Kuteva exclude the notion of reanalysis, they do not, in principle, object to the claim that reanalysis often co-occurs with grammaticalization. Hopper and Traugott claim that grammaticalization is always accompanied by reanalysis, but that not every case of reanalysis is a case of grammaticalization (2003 [1993]: 59). For instance, they claim that reanalysis may occur in types of change that include a shift from grammatical to lexical structure (e.g. autonomization of a bound clitic) and other changes that have their effect on the lexicon and not the grammar (e.g. compounding); hence, reanalysis is considered to be lexicalization instead (ibid. 58). However, some authors have found the notion of reanalysis to

be useful, at least in describing the development of complex adpositions (Hoffmann 2005). Likewise, in the present study, the mechanism of reanalysis is claimed to be one of the main mechanisms of change that contributes to the emergence of complex postpositions in Estonian. Thus, the present case of grammaticalization is claimed to occur hand-in-hand with reanalysis.

2.5.3.1. Reanalysis of complex postpositions

Langacker defines reanalysis "…as a change in the structure of an expression or class of expressions that does not involve any immediate or intrinsic modification of its surface manifestation" (Langacker 1977: 58). According to Harris and Campbell (1995: 61), reanalysis involves several types of change. The types that are relevant to the case of Estonian complex postpositions are: change in constituency, hierarchical structure, and category labels. The constituency of an expression indicates what goes with what (Hopper, Traugott 2003: 51). The change in constituency of Estonian *selja taga* 'behind one's back' (see example (41)) shows that *minu selja* ('my back') no longer forms a nominal phrase consisting of a genitive modifier (*minu*) and a head noun (*selja*), but instead *selja taga* (back+behind) forms a unified item, namely a complex postposition *selja+taga*. This also constitutes a change in the category labels – reinterpretation of the noun *selja* (back-GEN) as part of a complex postposition.

(41)   [[minu-Ø   selja-Ø]   taga]   >   [minu-Ø   [selja-Ø   taga]]

       I-GEN   back-GEN   behind   I-GEN   back-GEN   behind
       'behind my back'   'after him/her (temp.)', 'behind him/her',
                                                   'in his/her absence', etc.

Along with change in constituency comes a change in the hierarchical structure of the expression (as shown in (42)). As the reanalyzed *selja taga* forms a holistic item, a complex postposition, it also becomes the new head of the complex phrase. Before reanalysis, the simplex form *taga* ('behind') acted alone as the head of the postpositional phrase.

(42)   A. [[minu-Ø   selja-Ø]   taga]   >   B. [minu-Ø   [selja-Ø   taga]]

       NP   P   N   P
                                     PP   PP
       'behind my back'   'after him/her (temp.)', 'behind him/her',
                                                   'in his/her absence', etc.

The above examples are instances of syntactic reanalysis. However, reanalysis is not confined to (morpho)syntactic processes only. Reanalysis may also occur without grammaticalization. Hopper and Traugott (2003 [1993]: 59) have pointed out that sometimes reanalysis co-occurs with lexicalization. They illustrate the

claim with examples of formation of compound nouns, such as *hussy* < *house wife* that they claim to be instances of reanalysis because the boundary between the words that are compounded is lost. This process of reanalysis is considered to be lexicalization because the effect of this change (development of a compound) is on the lexicon. However, when complex postpositions evolve, it also affects the grammar.

It has been pointed out by many authors (e.g. Brinton,Traugott 2005; Eckardt 2006) that structural reinterpretation co-occurs with semantic changes. Eckardt (2006) has pointed out that the prominent role of semantic change in reanalysis suggests that reanalysis is triggered by semantic changes. It is likely that semantic change, i.e. lexicalization makes the simple postpositional phrases susceptible to reanalysis as complex forms. The lexicalization of the formerly freely combined unit is conducive to the unit interpretation of the phrase. The NP string is considered to be a more tightly bound unit in the utterance (as in (42-B)). Thus, in (42-B) *selja taga* (back+behind) is analyzed as a fixed unit within the utterance, and, therefore, the link between the former parts of the NP (as in (42-A) are weakening, whereas the bond between the new complex form is strengthening.

2.5.3.2. Actualization

Reanalysis is considered to be a covert process, i.e. reanalysis does not neces-sarily include any formal changes (Langacker 1977: 58). Hoffmann (2005: 57), too, points out that there is no formal difference between the freely combined PNP sequence and a grammaticalized complex preposition in English. Whether we analyze the phrase *in view of* as a freely combined phrase (*in view of the mountain*) or as a complex postposition (*in view of the facts*), its surface manifestation remains the same. However, while reanalysis is in itself invisible, it may bring about changes that are formally detectable and confirm that re-analysis has, indeed, occurred (Langacker 1977: 58). The process of emergence of formal evidence is often referred to as 'actualization'. Actualization takes place after reanalysis and is considered to be the formally detectable realization of reanalysis. In the following, I demonstrate the features that suggest that the body part related utterances have been reanalyzed as complex postpositions.

Firstly, the data suggests that the change mostly affects the postpositional phrases in the singular form, that is, the complex items prefer the form *selja-Ø + taga* (back-SG.GEN + behind) not *selga-de+taga* (back-PL.GEN + behind). The preference for the singular form indicates that the first component of the phrase, i.e. the body part noun, is losing its morphosyntactic properties, i.e. its ability to be pluralized. The preference of the complex function words for the singular form has brought about another phenomenon, namely that of non-agreement between the body part noun and the preceding noun (i.e. its 'former' modifier). We can observe in (43) that *liidrite* 'leaders' is plural, whereas its former head *selja* 'back' is singular. The non-agreement is considered to

manifest decategorialization because, as a regular lexical item, the noun would be expected to agree with its modifier.[21]

(43) *Eesti   on   uuendussuutlikkuselt   jõud-nud   Euroopa-Ø*
      Estonia   be.3SG   by innovation ability   reach-PST.PTCP Europe-GEN
      Liidu-Ø   __liidri-te__   __selja-Ø__   __taha__.
      Union-GEN   leader-PL.GEN   back-GEN   behind.LAT
      Lit. As for innovation, Estonia has gotten behind the back of European leaders.
      'As for innovation, Estonia has caught up with the leaders of the EU.'

This is taken to be indicative of decategorialization, a mechanism of grammatical change which has been described by many authors in connection with grammaticalization (e.g. Hopper, Traugott 2003), and by some (Heine, Kuteva 2002; 2007), it is regarded as one of the basic parameters of grammaticalization.

However, non-agreement alone cannot be taken to determine the status of a simple or a complex gram because non-agreement is not entirely impossible in lexical usages. For instance, non-agreement may occur in certain contexts where the meaning of the body part is more general (as in (44)), where the body part term stands for the human back in general. Non-agreement is also possible in other types of lexicalizations that do not lead to the development of a complex function word. These can be observed in (45) where the body part term *selg* 'back' is used in a semi-productive idiomatic expression *selga prügiseks tegema* 'to pin somebody down', and in (46), where the body part term functions as the simple postposition, *seljas* '(have) on'.

(44) __Meie-Ø__   __selg__   *väga   otseselt   peegelda-b   meie-Ø*
      our-GEN   back   very   directly   reflect-3SG   our-GEN
      psühholoogilis-t   seisundi-t.
      psychological-PRT   state-PRT
      'Our back reflects our physical state quite directly.'

(45) *Flora-Ø   poolkaitsja   tea-b   et   nen-de   __mees-te__*
      Flora-GEN   midfielder   know-3SG   that   these-PL.GEN man-PL.GEN
      __selg__   *prügise-ks   teh-a,   tule-b   veel   üksjagu   treeni-da.*
      back   dusty-TR   make-INF   must-3SG   yet   quite a bit train-INF
      Lit. The FC Flora midfielder knows that in order to get some dust on the back of these men, one must practice quite a bit.
      'The FC Flora midfielder knows that in order to pin down those men, one must practice quite a bit.'

---

[21] It should be noted that this is not a formal restriction to all NPs. Formally, the genitival modifier does not have to agree with its head noun. Thus, the utterances *poisi raamat* 'the boy's book' and *poiste raamat* 'the boys' book' are equally acceptable in Estonian. However, in order to make sense semantically, the agreement in number is necessary in NPs consisting of a complement noun and inalienably possessed object, such as body parts. It has to do with the fact that under normal conditions these cannot be shared, i.e. it is impossible for several persons to possess a mutual back.

(46) *Mantel   on   iga-Ø   aasta-Ø   must-have   kuid*
       overcoat   be.3SG   every-GEN year-GEN   must-have   but
       väikes-te   pois-te   seljas   väga   harva   näh-a, ...
       little-PL.GEN boy-PL.GEN on.LOC   very   rare   see-INF
       Lit. An overcoat is an every year must-have but not too often seen in the back of
       little boys.
       'An overcoat is an annual must-have, but it's not often seen on little boys.'

Another phenomenon that has been spotted in the use of the body part related postpositional phrases is extension beyond human reference. As lexical items, body parts are mostly used in connection with human referents, i.e. the first component of the postpositional phrase is usually preceded by a word that refers to a human being. The data suggests that in some contexts, body part terms are complemented by a noun that refers to a (human) collective or institution, or an abstract notion. The use of a linguistic item in new contexts has been described in connection with the grammaticalization process by many authors (*context expansion*, cf. Himmelmann 2004). In this study I have adopted the terminology of Heine and Kuteva (2002, 2007) who use the term *extension*. Extension is considered to be one of the basic parameters of grammaticalization (Heine, Kuteva 2007: 34). The extension in Estonian body part related postpositional phrases is observable in (47), where the word *käsi* 'hand' is preceded by *Jaguar*, which refers to the company rather than a human being.

(47) *Viimane   korralik   **Jaguari-Ø**   käeall*   *teh-tud*
       last   decent   Jaguar-GEN   hand.under.LOC   make-PST.PTCP
       jaguar   on   III   Seeria.
       Jaguar   be.3SG 3rd   series
       Lit.The last decent Jaguar made under the hand of Jaguar is from series III.
       'The last decent model made by Jaguar is the Series III.'

As with non-agreement, extension does not necessarily indicate grammaticalization in all cases. Like in many other languages, body parts term refer to object parts as well (e.g. *tooli selg* 'back of a chair'). While such cases manifest extension, they do not necessarily manifest the extension of the whole phrase (noun + postposition), but rather the extension of the body part term alone. The extension of the bare noun is not considered to comprise evidence of grammaticalization in cases where the phrase is still analyzable as being freely combined. The extension of the whole phrase is manifested in usages where the noun preceding the phrase refers to a non-individual or frontless object (as in (47)). These cases make it clear that it is the extension of the whole phrase, and can thus be considered as evidence of grammaticalization.

   The above discussion on extension primarily affects the semantic features of the noun functioning as the complement of the complex postposition. However, the extension of a linguistic item is also observable in larger syntactic contexts in which the emerging complex item occurs. For instance, the extension of a

grammaticalizing item is reflected by the abundance of other lexical items that it co-occurs with, i.e. how productive the element is in usage. In this study the productivity of the postpositions is observed amongst the lemmas of PNs and verbs co-occurring in the same clause with the body part related postpositional phrases (see section 2.5.4).

Non-agreement and extension are considered to be not only formal evidence of the occurrence of, but also reinforcement of, reanalysis. The more incompatible the first component becomes with the preceding element, the more closely bounded the new complex structure becomes.

2.5.3.3. Abruptness of reanalysis

Reanalysis is usually considered to be an abrupt change in the sense that a linguistic item is analyzed as belonging to one or the other category (Hopper and Traugott 2003 [1993]; Brinton and Traugott 2005, *inter alia*). This means that it may not carry features of the source and the target category at the same time. However, this claim has recently been called into question, mainly because it is in conflict with the gradual nature of language change in general and the process of actualization (Haspelmath 1998, De Smedt 2012). Haspelmath (2011: 345–346) argues against the common belief that reanalysis whereby syntactic phrases become complex words is an abrupt process. Instead, he shows that there is no solid way to distinguish between items pertaining to syntax and morphology. He goes so far as to claim that the distinction between syntax and morphology is highly doubtful, and it is not clear whether the two-way distinction (words vs phrases) of the syntax-morphology continuum is appropriate (Haspelmath 2011: 352). De Smedt (2012) has claimed that non-abruptness of reanalysis is manifested by the existence of so-called 'hybrid' forms. It can be observed in example (48) that there are usages of *lot of*, which exhibit features of both structures – noun and quantifier. The presence of adjective *awful* suggests *lot* behaves as a noun but the non-agreement suggests that it behaves as a quantifier (De Smedt 2012: 142).

(48)   *An awful lot of people are on medication who don't need it.* (De Smedt 2012: 142)

Similar cases can be found among Estonian body part related postpositional phrases. For instance, there are usages (as exemplified in (49)) that carry several features of complex postpositions: the usages are lexicalized, the first component of the body part related phrase does not agree with its preceding element, and the preceding element is also semantically incompatible with the body part noun.

(49) *Ometi   saa-b   tänase-Ø   tarkuse-Ø   pinnalt   öel-da,   et*
       though   can-3SG   today-GEN   wisdom-GEN   based   say-INF   that
       *Ameerika-Ø   **Ühendriiki-de**   **laia-Ø**   **selja-Ø**   **taga***
       America-GEN   united state-PL.GEN   broad-GEN   back-GEN   behind.LOC
       on   Euroopa-Ø   riigi-d   viimase-Ø   poolsajandi-ga
       be.3SG   Europe-GEN   country-PL   last-GEN   half a century-COM
       muutu-nud   enese-ga   petlikult   rahuloleva-te-ks.
       change-PST.PTCP   self-COM   deceitfully   satisfied-PL-TR
       Lit. Based on today's wisdom, it can be said that behind the broad back of the
       United States of America, European countries have become deceptively satisfied
       with themselves.
       'Based on what we know today, it can be said that behind the USA, European
       countries have become deceptively satisfied with themselves.'
       [www.diplomaatia.ee]

However, this example also includes an element which prohibits us from analysing it as a complex postposition – the adjective *laia* ('broad'). As stated above, adjectives are not considered suitable for reanalysis as complements of complex postpositions. Because the postposition and its complement have to be immediately adjacent (Palmeos 1985: 3, for Estonian; cf. Suutari 2006: 112 for Finnish; cf. van Pareren 2013: 94, for Mordvin), the proper noun *Ühendriiki-de* (United State-PL.GEN) cannot be analyzed as the complement of the complex postposition but rather as the adjectival complement of the noun *selja* (back). The latter analysis suggests that (49) is an example of a freely combined phrase. As mentioned above, the loss of ability of the nominal component to be pre-modified by an adjective is also considered characteristic of complex pre-positions in various languages (e.g. English (Hoffmann 2005: 56; Dutch Moirón, Bouma 2003: 157; and Swedish Sigurd 1993: 202).

2.5.4. Frequency, fixedness, and productivity

No study of grammaticalization that makes use of corpus data can escape addressing the role of frequency. Frequency is considered to play a significant role in most accounts of grammaticalization (e.g. Krug 2000; Hopper, Traugott (2003) [1993]; Bybee 2003, 2007, 2010; Hoffmann 2005). Grammaticalization is usually associated with high frequency. Due to the general meaning of grammatical items and their ability to occur in more diverse contexts, grammatical items are more frequent than lexical items. Thus, as grammaticalization by definition involves the development of grammatical items out of lexical items (or less grammatical items), an increase in frequency is eminent in the process of grammaticalization. (Bybee 2003: 602)

   Yet there is still some obscurity about the exact nature of the role frequency plays in the process. First of all, it must be noted that increase in frequency does not equal grammaticalization. For instance, Mair (2004: 125) claims that increased frequency does not always suggest grammaticalization but may be

related to other phenomena or extra-linguistic factors. Secondly, it has been suggested that high frequency is not vital for grammaticalization to occur. For instance, Hoffmann (2004; 2005), who has studied the grammatical status of low-frequency NPN-constructions (such as *in proximity to*) has suggested that such low-frequency items may be considered complex prepositions on the basis that they may be affected by analogy with more frequent, formally similar constructions. Moreover, he underlines that when studying frequencies, one must take into account that some concepts are expressed less frequently than others, and if the linguistic element under study is the preferred means of expressing a concept, it might be more salient than would be concluded based on its overall frequency (Hoffmann 2004: 204–205; Hoffmann 2005: 164). Moreover, there is still some dispute about the causal relationship between frequency and grammaticalization – it is not clear whether high frequency is the result or the prerequisite (or concomitant) of grammaticalization (Mair 2004: 126). Mair (2004) attempts to answer this question by analyzing several cases of grammaticalization, among them the development of the *going to*-future. He concludes that a rise in overall frequency is often a delayed result of grammaticalization which has occurred centuries earlier (Mair 2004: 38).

However, there are different ways to count frequency, and the results of analysis may be dependent on the method used. For instance, Mair (2004: 128–129) reports that despite a belated increase in frequency of *going to*, observing the frequencies of grammaticalizing items in relevant contexts yielded various results. For instance, his analysis showed that the uses of *going to* + INFINITIVE proportionally exceeded the contexts with prepositional complements already centuries before any change in overall frequency of the phrase took place (Mair 2004: 128–129). Thus, although grammaticalization may not be reflected in the overall frequency of the phrase, it might be observable as changes in proportion of usages that are relevant to the particular instance of grammaticalization. According to Bybee (2003: 604–605), frequencies may be observed as token frequency (text frequency) or type frequency (frequency of a pattern). Grammaticalization may be observed in both cases. Thus, in the present account, where the object of study occurs as the source form as well as the target form, it is useful to observe frequency in appropriate contexts.

Mair (2004: 123) also suggests that in regard to frequency, it is useful to distinguish two types of grammaticalization – 'dynamic' and 'static'. The former stands for the diachronic process observable in major shifts in frequency, and the latter for synchronic variation, whereby lexical items are occasionally used in a grammatical function. Instances of static grammaticalization are usually not associated with high frequency, nor are they usually detectable diachronically as a directed change. Mair suggests that such instances of grammaticalization may be better studied qualitatively (Mair 2004: 138–139). It must be noted that the present grammaticalization process – the development of complex postpositions in Estonian – is rather an instance of the static type. Thus, based on Mair, the corpus analysis cannot be expected to yield any drastic changes regarding the frequency of the phrases under investigation. Nevertheless, I will attempt to

account for frequency when tracing the development of complex function words diachronically (see section 4.8.).

As the diachronic data are few (see section 3.2.2), the analysis of complex function words focuses on synchronic data, where the *change* in frequency cannot be observed. However, in a synchronic study, pattern frequency can still be observed, especially in the context of the parameters of grammaticalization described above. In addition to raw frequencies of the phrases, I will observe the proportion of freely combined phrases and complex units, the proportion of adverbial uses and postpositional uses, and the frequency of uses that indicate actualization of reanalysis (contextual expansion and non-agreement).

In addition to observing absolute (or relative) frequencies, it is useful to implement statistical methods that measure associations between words. Association measures show the strength of association between the words (Evert 2005: 75). Association measures have advantages over absolute frequency measures because they allow us to determine whether there is a statistical association between the words or their co-occurrence is mere chance. For instance, two words that are both highly frequent may co-occur by coincidence, but association measures give a statistical interpretation of the relationship between the words (Evert 2005: 20–21). In this study, association measures are used for two purposes: measuring the collocational strength between the components of the phrases and measuring the collocational strength between the phrase and other elements in the sentential context.

The strength between the components of the phrase shows how tightly bound the units are. Tight connection between the components of (complex) structures is associated with increasing autonomy (Bybee 2010: 50), fixedness, freezing or fossilization, which have been associated with grammaticalization as well as lexicalization (Brinton, Traugott 2005: 105).[22] Here, fixedness is measured with mutual information (Church, Hanks 1990). The method is also used by Móiron and Bouma (2003) to measure associational strength in Dutch collocational prepositional phrases. Mutual information compares the probabilities of occurrence of a phrase to probabilities of occurrence of each component of the phrase independently. If the co-occurrence of a body part term and a simple postposition (such as *kaela peal* (neck+on)) is not due to chance, the mutual information of the components (*I*) is above 0. (Church, Hanks 1990: 77) In order to demonstrate that the scores, indeed, suggest fixedness, the values of mutual information of the phrases in question will be compared to those of body part related phrases that do not behave as complex units (*selja taga* (back+behind) vs. *pea all* (head+under)), as well as to body part related phrases that consist of the same components as the phrases under investigation, but are formed with plural body part terms (*selja taga* (back+behind) vs. *selgade taga* (backs+behind)) (see section 4.1.).

---

[22] To some extent, these terms are (e.g. Brinton, Traugott 2005) used as synonyms. To avoid confusion, henceforth, only *fixedness* will be used to refer to the fixation of the studied phrases.

The strength between the phrase as a whole and other elements in the sentential context is used to observe the productivity of the complex units. Productivity is here understood as the ability of a linguistic item to be used repeatedly to produce more instances of the same pattern (Crystal 2000: 310). Following Brinton and Traugott, productivity is here taken to be a scalar notion, i.e. there are are more productive and less productive linguistic items (Brinton and Traugott 2005: 18). While grammaticalization often starts out in narrow contexts, the process of grammaticalization is associated with increase in productivity and grammatical items are considered to be of high productivity. (Brinton, Traugott 2005: 17–18, 100, 109). Thus, the productivity of the studied phrases is of interest here because it can be used as one the factors to determine the degree of grammaticalization among the studied phrases.

As the development of the function words studied here is still in its initial stages, it may be assumed that their use is still partially contextually restricted, i.e. unproductive. Indeed, based on dictionaries, most of the studied phrases have been treated as instances of figurative language. Most of the studied phrases are listed in the Phraseological Dictionary[23] (Õim 2000) either as separate entries or as a part of a larger fixed expression. For instance *selja taga seisma* lit. 'stand behind [one's] back' has been listed as a phraseological expression meaning 'to support somebody'. The same dictionary does not list *käekõrval* as part of any fixed expression. However, the he database of Estonian verbal multi-word expressions[24] lists *käekõrvale võtma* 'take [something] beside [one's] hand' as a multiword expression[25]. The aim of the analysis of productivity in the present study is to systematically determine:

i   strong collocates, which are suggestive of formulaic use of the studied phrases;
ii   the amount of examples that represent such formulaic uses and amount of examples that are freely combined.

Rich contexts are considered to suggest productive use of the complex units. However, if the use of the complex units is confined to certain restricted contexts, they may not be considered as grammatical items but rather as instances of fixed expressions.

In this type of grammaticalization, the productivity of the complex items is observed in two aspects – the occurrence of the complex postposition with a (pro)nominal complement (e.g. *euro* in example (50)) and the verb that co-occurs with the complex item. The verb that co-occurs with the complex unit is the verb that, together with the body part related complex item, expresses the relationship between the LM and the TR (*varitsema* 'ambush' in (50)).

---

(50) *Kuus aasta-t ELi-s ei ole-ø eestlas-t kuigivõrd*
six year-PRT EU-INE NEG be-CONNEG Estonian-PRT much
muut-nud, sest kolm kuu-d enne Euroopa-ø
change-PST.PTCP because three month-PRT before Europe-GEN
ühisraha-le ülemineku-t on palju-d asu-nud
common currency-ALL transition-PRT be.3PL many-PL start-PST.PTCP
euro-ø selja-ø taga varitse-va-Ø *hinnatõusu-ø*
Euro-GEN back-GEN behind.LOC lurk-PTCP-GEN price rise-GEN
hirmu-s oma-ø sääst-e kuluta-ma. [www.maaleht.ee]
fear-INE own-GEN saving-PL.PRT spend-SUP
'Six years in the EU has not changed the Estonian much because three months
before the conversion to the European common currency many have started to
spend their savings in fear of the price rise lurking behind euro's back.'

To determine the association of the complex items and the (pro)nominal
complement and the verb, a log-likelihood measure is used. The log-likelihood
measure takes into account the frequency of both linguistic elements, the
frequency of their co-occurrence, and the size of the corpus. The higher the log-
likelihood score, the more closely bound the word pair. Log-likelihood is a
widely used measure in linguistics. It can be used to find idiomatic expressions
or other fixed word combinations and formulaic expressions. (See Evert 2005:
21). This measure has been used on Estonian data (Uiboaed 2010), as well as to
determine the strongest collocates of English complex prepositions. For
instance, Hoffmann (2005: 78–79) implements this method to observe the
strongest verb collocates of the complex preposition candidate *in need of.* He
concludes that the very short list of collocates and very high association score of
the verb *be* suggests (along with other factors) that *in need of* should perhaps
not be included in the list of common complex prepositions.

III MATERIAL AND DATA SOURCES

This chapter is concerned with the data analyzed in the present study. In the following I will describe the selection of the postpositional phrases studied here, give an overview of the data sources and explain their selection.

3.1. The object of study

In this study, the development of complex function words is observed in a small group of body part related postpositional phrases (body part term + spatial postposition). In previous research, other phrases have been investigated, according to the purpose of the studies. For instance, Habicht and Penjam (2007) hypothesised that the status of complex function word (which is indicated by spelling) is associated with the degree of grammaticalization of the simple postposition as well as part of speech (pronoun vs. noun) and length of the complement. Jürine (2011) hypothesised that the spelling of the phrase is connected with literal vs. non-literal interpretation of the phrase. Accordingly, the phrases examined were chosen based on these variables and intuition. However, intuitively it seems that the development of complex function words (especially adverbs) is quite a widespread tendency. Thus, in this study, I investigate whether (and to what extent) this process is attested in the body part related postpositional phrases. The body part related phrases are an interesting object of study for the following reasons:

i.   In the world's languages, body parts are a frequent source of function words, especially adverbs and adpositions (c.f. Svorou 1994; Heine 1997, 1989; Heine, Kuteva 2002, 2007, etc). The same process has been shown to have occurred in the case of simple function words in Estonian and Finnish (Habicht 2001, Ojutkangas 2001). Thus, it is probable that body part related phrases are a potential source of complex function words as well.

ii. Body part related postpositional phrases are a semantically homogenous group of expressions. As such, they provide a uniform set of examples. This enables me to draw parallels between them and generalize the results, whereas studying semantically and functionally different expressions might not untangle the development of complex function words in such detail.

iii. A simple observational inspection of the body part related postpositional phrases uncovers that some of them (see below) are already presented in dictionaries as compound adverbs.[26] Thus, they are likely to be used in contexts where they may be analyzed as complex postpositions as well. For

---

[26] The term 'compound adverbs' suggests that they are compounds, i.e. recognized as holistic units by language planning, and, therefore, written as a single word. As mentioned earlier (see section 2.3), in this study I use the broader term 'complex adverbs' (and postpositions) which are defined based on semantic and syntactic features of the phrase.

instance, Karelson (2005: 65–66), who discusses problems with word class determination based on the data obtained for compiling 'The Estonian Explanatory Dictionary', claims that the compounds *seljataga* (back+ behind) and *käekõrval* (hand+beside) are used as adverbs as well as postpositions and listed as such in the dictionary as well. However, having separate entries in the dictionary is not, of course, taken as a criterion for the selection of the phrases.

iv. Body part related phrases are an adequate source for the description and determination of the complex function words because they provide a sample of grammaticalizing items that are present in the contemporary language as the source form (simple postpositional phrase) as well as the target form (the complex adverb, the complex postposition). Thus, they enable me to characterize the use of each expression as the simple and the complex structure and, thereby, determine and test the criteria for distinguishing the complex items from the freely combined phrases.

It follows from the final reason listed above that one criterion for the selection of a phrase is that it also allows a literal interpretation. Therefore, the second component of the phrase is always a spatial postposition (e.g. *selja taga* (back+behind) 'behind one's back'). The literal interpretations of the phrases are easily contrasted with unit meanings (i.e. lexicalized meanings), which are usually more abstract. Thus, the two concrete components ensure that when the phrase expresses a more abstract meaning, it is due to lexicalization of the phrase, and not to the abstractness of the simple gram. For instance, the body part term *käsi* 'hand' and a causal postposition *läbi* 'through'[27] forms a postpositional phrase which carries an abstract meaning (as in example (51)) and appears in contexts were analysis as a complex adverb or a complex postposition is possible.

(51) *Kui   kogu   Starki-Ø   isiklik   maailm   verivaenlase-Ø*
      when   entire   Stark-GEN   personal   world   arch enemy-GEN
      käe-Ø   läbi   hävita-ta-kse,   asu-b   ta   piina-Ø
      hand-GEN   through   demolish-IMPS-PR   set-3SG   s/he   torture-GEN
      ja   seiklusrikka-le   otsingu-le   et   süüdlase-d   üles   lei-da.[28]
      and   adventurous-ALL   journey-ALL   to   responsible-PL   up   find-INF
      Lit. When Stark's world is demolished through the hand of his arch enemy, he begins an agonizing yet adventurous journey to find those that are responsible.
      'When Stark's world is demolished by his arch enemy, he begins an agonizing yet adventurous journey to find those that are responsible.'

---

[27] The adposition *läbi* 'through' does express a spatial meaning as well. However, here it clearly expresses CAUSE because as a spatial adposition it either is preposed (*läbi akna-Ø* (through + window-GEN)) or takes a complement in the elative, not the genitive, case (*akna-st läbi* (window-ELA through), *käe-st* läbi (hand-ELA through)).
[28] http://www.forumcinemas.ee/Event/298965/ (Accessed 11.01.2016)

It may be argued that in (51), *käe läbi* (hand+through) is lexicalized and expresses causality. However, as the simple postposition itself carries causal meaning, it may just as well be argued that the abstract sense may be formed by freely combining the two components. In this case, there would be no phrasal meaning and no lexicalization. Therefore, it would not make sense to talk about complex grams, but rather postpositional phrases that express abstract meaning(s). Thus, to minimalize the vagueness of the semantic criteria and to unequivocally distinguish the simple and complex structures, only post-positional phrases with spatial grams were investigated.

This study analyzes the following body part related phrases: ***selja taga*** (back+behind), ***käe all*** (hand+under), ***käe kõrval*** (hand+beside), ***külje all*** (side+under), ***kaela peal*** (neck+on), ***jalge all*** (feet+under). This list is not exhaustive, and it is not meant to be understood to represent the only selected phrases which bear the relevant criteria. The selection of the phrases was executed as follows.

I listed all the potentially relevant body part related phrases by combining a list of body part terms with a list of simple postpositions expressing spatial functions. The list of body part terms is based on frequent sources of grams in various studies (Svorou 1994; Heine, Kuteva 2002). The list includes basic terms only, i.e. it does not contain complex terms (e.g. *küünar+nukk* 'elbow') or derived forms (e.g. *ist-mik* 'buttocks'). The terms for internal organs as well as obscene vocabulary were excluded because these are not likely to undergo grammaticalization for obvious reasons. The list of simple grams expressing spatial meanings is adapted from Palmeos (1985). In order to provide a more uniform sample, only the postpositions that have three locative forms (the lative, locative, and separative) were included (see section 2.2.).

The potential phrases were combined of 35 body part terms and 24 spatial grams (see Annex 1), amounting to a total of 840 postpositional phrases. Most of these phrases express compositional meanings. Such usages are always ana-lyzed as simple structures and, given the aim of this study, are not of particular interest. The following criteria were implemented to determine potential complex units:

i.   the phrase is lexicalized, i.e. carries a meaning that is not directly derivable from its components;
ii. the phrase is not confined to extremely restricted contexts in contemporary language;
iii. the phrase occurs in contexts that allow structural reanalysis as a complex postposition;
iv. the semantic shift is clear enough to distinguish simple and complex struc-ture;
v.   lexicalized usages make up a considerable amount of data, i.e. enough to analyze the complex unit.

3.2. Data sources

In the present study, synchronic as well as diachronic analysis is implemented. The synchronic analysis is based on data extracted from etTenTen Corpus. The diachronic analysis is based on data that was obtained from the Corpus of Old Literary Estonian (COLE), the Corpus of 19th Century texts, and the Corpus of Estonian Literary Language (CELL). In the following, I will explain the selection of the data sources and give a brief overview of these corpora.

3.2.1. Synchronic data

The synchronic analysis of the development of complex function words is based on data that has been obtained from etTenTen – the Estonian Internet corpus. Metslang (2006: 190–191) has also pointed out that the web is a prolific source of material for linguistic research and that the non-edited texts that prevail on the Internet might turn out to be an accelerator of language change. The web is considered the best data source for this study in particular, because the language material available online represents contemporary language. It represents natural language use which is not necessarily constrained by the rules of standard language, while still bearing features of written language. Crystal suggests that although the Internet represents a 'mixed medium', Internet language should be considered to be written language with traits of spoken language, not the other way around (Crystal 2011: 21). As the development of complex function words, especially complex postpositions, is considered to be in its early stages, and has probably not yet been established in standard written Estonian, the Internet data is considered to be the most appropriate data source.

etTenTen is an Estonian Internet corpus with a size of 270,000,000 words.[29] etTenTen belongs to the etTenTen family (c.f. Jakubíček 2013). The corpus has been developed in cooperation between Lexical Computing Ltd., Filosoft LLC, and The Institute of Estonian language (Kallas et al. 2015). The texts are morphologically and syntactically analyzed and automatically tagged. The data has been collected from 686,000 websites and the texts represent the following categories: government (2%), forum (20%), religion (3%), blogs (10%), periodicals (25%), informative (7%), and unknown (unclassified) (32%). Thus, the corpus includes texts that represent language use that is (close to) standard written language (*government*, *periodicals*, to a certain extent *informative*) as well as texts that are possibly closer to spoken language and less likely to adhere to the rules of standard language (e.g. forum, blogs).[30] Therefore, it suits the purpose of this study. etTenTen is also considered to be a suitable data source for the present study because of its size. In order to be able to extract a

---

[29] The corpus is available at www.keeleveeb.ee and downloadable at
http://downloads.sketchengine.co.uk/ettenten13.processed.prevert.xz
[30] http://www2.keeleveeb.ee/dict/corpus/ettenten/about.html (Accessed 11.01.2016)

sufficient amount of data on the bi-grams in the study, the corpus has to large enough. The 270,000,000 word corpus is the largest corpus of Estonian.

3.2.2. Diachronic data

Although the main focus of this study is on synchronic analysis, diachronic data has been considered as well. The main reason for focusing on synchronic variation is the relatively recent nature of the development. However, as grammaticalization is a gradual change it should be best observed diachronically. Nevertheless, before I introduce the data sources of diachronic study, there are some methodological considerations that need to be acknowledged.

The main issue is the lack of diachronic data. The literary tradition of Estonian does not go back more than five hundred years. The oldest Estonian texts available electronically are from the 16th century. Also, due to the relatively small size of the diachronic corpus and the structural complexity of the object of study, the search yields few results. This sets restrictions on the conclusions to be drawn from the diachronic data. Another potential problem concerns the compatibility of the available corpora. In order to observe the development of body part related postpositional phrases over the centuries, I used three corpora: the Corpus of Old Literary Estonian (COLE),[31] which covers the period from the 16th century until the end of the 18th century, the Corpus of 19th Century Texts[32], and the Corpus of Estonian Literary Language (CELL)[33], which covers the period from the 1890s until the 1990s. A potential problem with compatibility arises from the fact that the corpora are compiled of texts of different genres. The corpora are more thoroughly introduced below. Here, suffice it to mention that COLE is mainly compiled of religious texts, the corpus of 19th century texts consists of religious texts as well as journalistic texts and fiction, and CELL consists of journalistic texts and fiction. Another potential issue arises from the fact that the corpora differ in size as well as the length of periods they cover. Although all of the frequencies in this study will be presented as relative frequencies (occurrences per million words), data from corpora that cover periods of vastly different length (such as CELL and COLE) cannot be examined side by side without criticism.

Despite the caveats mentioned above, these corpora remain the only electronically accessible collection of texts representing Written Estonian from the period of Old Literary Estonian to the 19th century, and through the 20th century. Thus, the corpora provide a valuable body of texts that reflect how the language has changed through that time. Therefore, COLE, 19th century texts, and CELL are also useful sources of data if one's aim is to describe the beginnings of a fairly recent development, such as the development of complex grams.

---

[31] http://www.murre.ut.ee/vakkur/Korpused/Kwic2/paring.htm (Accessed 02.01.2016)
[32] http://www.murre.ut.ee/vakkur/Korpused/Kwic2/paring19.htm (Accessed 02.01.2016)
[33] http://www.cl.ut.ee/korpused/kasutajaliides/index.php?lang=en (Accessed 02.01.2016)

3.2.3. The corpora used for the diachronic study

The Corpus of Old Literary Estonian (COLE) (size 1,550,802) mostly comprises Estonian texts but includes some parallel text in German as well. In terms of genre, religious texts prevail (80% of the texts in the corpus), while the rest of the texts represent non-fiction, fiction and journalism.

Most of the texts in COLE are morphologically tagged. There is much orthographic variation in the texts of COLE. Thus, the search was made by lemmas. In order to gather data from the texts that are not morphologically tagged, I repeated the search using a simple string search. In order to cover all possible orthographical variants of the phrases, I used dictionaries of Old Written Estonian (Ehasalu et al. 1997; Habicht et al. 2000; Kikas 2002; Kingisepp et al. 2010) which list the possible orthographical variants of the above-mentioned body part terms.

The corpus of 19th-century texts comprises texts originating in the 19th century. The size of the corpus is 520,307. As to genre, the texts are mostly fiction and various sorts of educational texts. As the corpus is not morphologically tagged, the query tool allows a simple search (by string) only. Thus, here, too, I searched for each body part term separately, taking into account orthographical variation. However, in the 19th century, there is less orthographical variation than in the texts of COLE.

The corpus of Estonian Literary Language (CELL) comprises texts from the 1890s to the 1990s. The Corpus consists of several subcorpora. The subcorpora that were chosen for this study are the following – the corpus of the 1890s (348,000), the corpus of the 1900s (236,000), the corpus of the 1910s (418,500), the corpus of the 1930s (369,000), the corpus of the 1950s (308,000), the corpus of the 1960s (333,000), the corpus of the 1970s (425,000), the corpus of the 1980s (425,000), and the corpus of the 1990s (1,467,000). Adding up the tokens of these subcorpora, the size of CELL is 4,488,500 words. The texts that make up CELL belong to two genres – journalism (53%) and fiction (47%).

CELL is not morphologically tagged, and the query had to be made by string. I searched for each phrase separately. To enable searching for all three forms (the lative, locative and separative), I used regular expressions. For instance, for the phrase *selja taga,* the regular expression which finds the phrase in all the forms and every position in the sentence is *[Ss]elja ta[gh]an*t*.* Because of variation in spelling, the search was repeated with regular expression that find such instances too (e.g. [Ss]eljata[gh]an*t*). All the results from these corpora were included in the analysis.

IV RESULTS OF THE CORPUS STUDY

4.1. Frequency and fixedness of the studied phrases

This section covers the frequency and fixedness of the studied phrases. Fixedness is observed in the collocational strength between the components of the studied phrases compared to other, i.e. freely combined body part related phrases. It will be demonstrated that the studied phrases tend to occur with a singular body part term. Absolute frequencies of the studied phrases in the etTenTen corpus are shown in Figure 3.
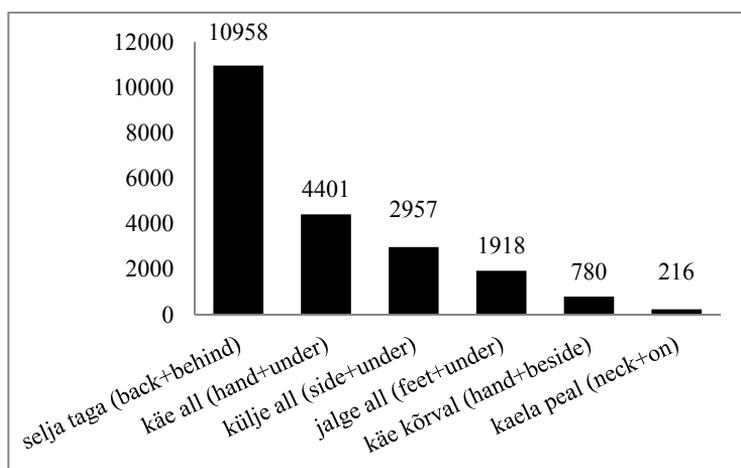
Figure 3. Absolute frequencies of the studied phrases in the etTenTen corpus (N = 270,000,000)

Figure 3 shows that the studied phrases vary in terms of their absolute frequency. The corpus of 270,000,000 words (etTenTen) contains as much as 10,958 instances of the phrase *selja taga* (back+behind), 4,401 instances of *käe all* (hand+under), 2,957 instances of *külje all* (side+under), and 1918 instances of *jalge all* (feet+under), while the phrases *käe kõrval* (hand+beside) and *kaela peal* (neck+on) occurred only on 780 and 216 instances respectively.

As grammaticalization is generally associated with high frequency (e.g. Krug 2000; Hopper, Traugott (2003) [1993], Bybee 2003, 2006, 2007, 2010; Hoffmann 2005), it could be concluded that the phrases of high frequency (e.g. *selja taga* (back+behind)) are further down the grammaticalization path. However, high frequency alone may not necessarily suggest grammaticalization. What is more, some authors (Mair 2004; Hoffmann 2004, 2005) have argued that high frequency is not a vital factor in grammaticalization. That is why the frequency should be investigated within the context of other relevant factors. First, I will discuss the frequency of the phrases as an indicator of fixedness. Fixedness is considered to be indicated by the collocational strength between

the components of the phrase (the body part terms and the simple postpositions) as well as the ability of the components of the phrase to display morphological variation. In the following, these indicators will be discussed one by one.

The collocational strength is measured using the mutual information measure (Church, Hanks 1990), which allows me to account for the absolute frequency of the phrases, the frequencies of the components of the phrases and the size of the corpus. Table 1 lists the mutual information (MI) scores of the studied phrases. $f(x,y)$ is the absolute frequency of the phrase, $f(x)$ the absolute frequency of the body part term and $f(y)$ the absolute frequency of the simple postposition in the etTenTen corpus.

Table 1. The mutual information score of the studied phrases in the etTenTen corpus (N = 270,000,000)

Phrase	MI	$f(x,y)$	$f(x)$	$f(y)$
selja taga (back+behind)	9.2	10,958	42,783	118,392
külje all (side+under)	6.1	2,957	37,077	326,702
jalge all (feet+under)	4.7	1,918	62,381	326,702
käe all (hand+under)	4.3	4,401	186,068	326,702
käe kõrval (hand+beside)	3.5	780	186,068	100,448
kaela peal (neck+on)	2.9	216	17,690	445,669

The mutual information score indicates how likely is the occurrence of each of the phrases, i.e. the combination of their components. For instance, the MI score for *selja taga* (back+behind) of 9.2 indicates that the occurrence of this phrase is 512 (that is 2^9) times larger than would be expected by chance (Church, Hanks 1990: 25). Therefore, the data suggests that *selja taga* (back+behind) which is also the most frequent phrase in the etTenTen corpus exhibits the strongest intra-phrase association among the studied phrases.

Table 1 shows that the ranking of the phrases based on MI score does not always coincide with the ranking based on the absolute frequency of the phrases (cf. Figure 3). Although according to the MI score (9.2) *selja taga* (back+behind) is clearly the most strongly associated phrase, there are minor differences in the ranking of the other phrases. For instance, although the absolute frequency of the phrase *käe all* (hand+under) is higher than that of *külje all* (side+under), the latter has a higher association score (6.1) than the former (4.3). This means that the association between the components of the phrase *külje all* (side+under) is stronger than in the case of *käe all* (hand+under). Consequently, *külje all* (side+under) is considered to be a more tightly bound unit. According to Church and Hanks (1990), MI score of 6.2 suggests that the probability of the occurrence of the phrase is 64 (2^6) times higher than a coincidence, which they consider to be a relatively strong association (Church, Hanks 1990: 25). According to the MI score, the occurrence of *käe all* (hand+under) and *jalge all*

(feet+under) is 16 times more likely than by chance, which also suggests that both of them are strongly associated units. The same applies for the less frequent and less strongly associated *käe kõrval* (hand+beside) and *kaela peal* (neck+on), which occurred approximately 8 times more often than is likely by chance. Thus, the MI scores of all of the studied phrases indicate that the occurrences of these phrases is not simply due to the high frequency of the components, but also owing to the strong to very strong association between the components of each phrase.

Of course, a strong association need suggest neither grammaticalization nor lexicalization. Often, it merely shows the company a word keeps. Some adpositional phrases may be more or less fixed, not because they form a holistic unit, but because certain nouns prefer certain adpositions (e.g. *puu otsa* (tree+aloft) not *puu sisse* (tree+in) 'up a tree'). Thus, it may seem that the occurrence of any plausible adpositional phrase (e.g. *käe kohal* 'above one's hand') is likely to occur more often than by chance. In order to demonstrate the fixedness of the studied phrases, they are contrasted with eight freely combined postpositional phrases combined using body part terms and locative postpositions (see Table 2).

Table 2. Mutual information and absolute frequencies of freely combined phrases in the etTenTen corpus (N = 270,000,000)

Phrase	MI	f(x,y)	f(x)	f(y)
nina kõrval (nose+beside)	1.6	19	17,108	100,448
pea all (head+under)	1.3	338	112,983	326,702
kaela all (neck+under)	0.7	34	17,690	326,702
selja all (back+under)	−0.2	46	42,783	326,702
külje taga (side+behind)	−1.4	6	37,077	118,392
jalge peal (feet+on)	−2.7	16	62,381	445,669
käe kohal (hand+above)	−3.7	7	186,068	130,673
käe taga (hand+behind)	−5.4	2	186,068	118,392

Tables 1 and 2 indicate that the absolute frequencies and MI scores of the freely combined phrases are considerably lower than those of the studied phrases. Although the absolute frequency of *pea all* (head+under) is a little higher (338 instances) than in case of other phrases in Table 2, all of the MI scores remain rather close to zero. The negative values indicate negative association between the components, i.e. that the words rather tend to avoid each other's company (Evert 2008). Thus, despite of consisting of components semantically close to the studied phrases and bearing plausible compositional meanings, the data does not suggest that the phrases presented in Table 2 are strongly associated or fixed.

Another indicator of the fixedness of the studied phrases is taken to be their ability to display morphological variation. As postpositional phrases are rather structurally restricted (N_{GEN} + PostP), the only possible variable is the gram-

matical number of the noun. It will be demonstrated that all phrases (except for *jalge all* (feet+under)) have been fixed in the singular form and barely occur in the plural at all. The absolute frequencies of the singular (f_{SG}(x,y) and plural nouns (f_{PL}(x,y)) and the MI scores are shown in Table 3.

Table 3. Postpositional phrases with singular and plural nouns in the etTenTen corpus (N = 270,000,000)

Phrase	f_{SG}(x,y)	MI (SG)	f_{PL}(x,y)	MI (PL)
selja taga (back+behind)	10,958	9.2	4	NA[1]
käe all (hand+under)	4,401	4.3	124	–0.9
külje all (side+under)	2,957	6.1	10	–2.2
käe kõrval (hand+beside)	780	3.5	1	NA
jalge all (feet+under)	243	1.7	1,918[2]	4.7
kaela peal (neck+on)	216	2.9	0	NA

[1] NA = not applicable. If absolute frequency is less than 5, an MI score cannot be calculated.
[2] The number of the exact phrase *jalge all*. Actually the number of plural form of this phrase is even larger because the numbers presented here only accounts for the phrases formed with the short plural genitive (*jalge*), and does not account for the phrases combined with the longer genitive (*jalgade*) which occur on 314 instances (MI = 2.0).

Table 3 shows that the studied phrases are generally less commonly used with plural nouns than with singular nouns. For example, *selja taga* (back+behind) occurs 10,958 times with the singular noun, but only occurs on 4 occasions in the plural. *Kaela peal* (neck+on) does not occur in the plural at all, and the queries for *külgede all* (sides+under) and *käte kõrval* (hands+beside) gave only a few results. With an absolute frequency of 124, *käte all* (hands+under) is the second most frequent phrase. Given that hands are body parts that normally come in pairs, this is quite expected. However, the negative MI score (–0.9) indicates that *käte all* is not a strongly associated phrase. Thus, it was observed that, in Estonian, the plural forms are, indeed, infrequent but still possible. Sigurd (1993: 199) lists the inability to pluralize the nominal component of the phrase as one of the criteria for 'the multi-word' prepositions in Swedish. The fact that the Estonian phrases with plural body part terms are not as strongly associated could be taken as evidence of greater fixedness of the combinations of singular body part phrase and simple postposition. However, the fact that the plural forms are rare may also be due to the so-called 'singular plural'(Alvre 1989: 68), an old characteristic of Estonian (but also attested in Finnish), which allows to refer to plural body parts formally in singular. This trait has receded but is preserved in some idiomatic expression (Õim, Õim 2015).

As the only phrase that has fossilized in the plural form, *jalge all* (feet+under) is more frequent in the plural (1918 instances) than the singular (243 instances). The MI score also indicates that the plural form *jalge all*

(feet+under) has a stronger association with *all* (under) (MI = 4.7) than its singular counterpart *jala all* (foot+under) (MI = 1.7).

In conclusion, the data suggests that the studied phrases are of different absolute frequency. The most frequent phrase *selja taga* (back+behind) occurs over 10,000 times, and the least frequent phrases *käe kõrval* (hand+beside) and *kaela peal* (neck+on) less than one thousand times. Nevertheless, the MI analysis indicates that all of the studied phrases are more or less internally strongly associated, i.e. their occurrence is likely due to more than simple chance. At the same time, the analysis of random but semantically plausible combinations of body part terms and simple postpositions did not yield scores suggestive of strong correlations. Moreover, the analysis suggests that the studied phrases prefer a single noun form, except for *jalge all* (feet+under) which prefers the plural. This suggests that as phrases become complex units, they also become more fixed. This is characteristic of grammaticalization (Bybee 2006: 715) and lexicalization (Brinton, Traugott 2005). In both cases, the body part noun exits the noun category in a certain fixed form, which is part of its process of becoming a fixed unit.

4.2. The distribution of freely combined phrases and complex units

So far, we have viewed only the formal aspects of the studied phrases and neglected to account for meaning. However, the development of complex function words is heavily dependent on their semantics. The body part related phrases that are studied here currently allow several interpretations. They may be interpreted as freely combined units (i.e. the simple structure) and complex units (i.e. the complex structure). Additionally, the data includes examples where the body part related phrase bares characteristics of both structures. Such usages are referred to as hybrid forms[34]. In the following, these structures will be characterized more thoroughly.

The **simple structures** (or **freely combined phrases**) consist of a body part noun and a simple postposition (52) or an object part noun and a simple postposition (53). The rare cases where the phrase is used in a figurative sense but is not lexicalized (54) are also considered to belong to the simple structure.

(52) *Abikaasa-ø*   **selja-ø**   **taha**   *astu-nud*   *Eva*
      husband-GEN   back-GEN   behind.LAT   step-PST.PTCP   Eva
      põimi-b   *käe-d*   *ümber*   *mehe-ø*   *kaela-ø*.
      entwine-3SG   hand-PL   around   man-GEN   neck-GEN
      'Eva, having stepped behind her husband's back, entwines her hands around her man's neck.' [www.naisteleht.ee]

---

[34] The term is used as in De Smedt (2012: 141), where it refers to linguistic forms that have characteristics of two underlying structures.

(53) *Turumaja-Ø*   **selja-ø**   **taga**   *pingi*   *peal*
market house-GEN   back-GEN   behind-LOC   bench-GEN   on-LOC
tuulavad   *Leida*   *ja*   *Silvi*   *paluka-i-d.*
winnow-3PL   Leida   and   Silvi   cowberry-PL-PRT
Lit. Behind the back of the market building on a bench, Leida and Silvi are rummaging lingonberries.
'Behind the market building on a bench, Leida and Silvi are rummaging ligonberries.' [www.vorumaateataja.ee]

(54) *Ta*   *reisi-b*   *ja*   *uuri-b*   *ning*   *püüa-b*   *leid-a*
s/he   travel-3SG   and   study-3SG   and   try-3SG   find-INF
vastuse-i-d   *väga*   *erisugus-te-le*   *küsimus-te-le –*   *kuidas*
answer-PL-PRT   very   different-PL-ALL   question-PL-ALL   how
toimi-b   *armastus,*   *kuidas*   *kiiga-ta*   *elu-Ø*   **seljataha,**
work-3SG   love   how   glance-INF   life-GENback.   behind.LAT
kuidas   *ava-da*   *us-t*   *igavikku-Ø?*
how   open-INF   door-PRT   eternity-ILL
Lit. S/he travels and explores and tries to answer various questions like – how does love work, how to peek behind the back of life, how to open the door to eternity?
'S/he travels and explores and tries to answer various questions like – how does love work, how to peek into the afterlife, how to open the door to eternity?' [www.saaremaa.ee]

It was discussed in section 2.5.1 that in order to be able to become a complex item, a postpositional phrase must develop a new meaning – one that is not directly derivable from the meaning of its components. Thus, the **complex structures** (or **complex units**) include all the usages where the body part related phrase is lexicalized. The same criterion is traditionally applied when determining the complex adverbs in the Estonian language (see section 2.2). The complex units may be realized as complex postpositions (55) or as complex adverbs (56) (see also section 4.4). As complex units, the body part related phrases usually carry more abstract meanings than as freely combined units (compare examples (52) and (55)), whereas there is not necessarily such a difference between the complex adverbs and complex postpositions (c.f. (55) and (56)).

(55) *Eriti*   *hea*   *on*   *puge-da*   *Friedmani-ø*   **selja-ø**
Especially   good   be.3SG   creep-INF   Friedman-GEN   back-GEN
taha,   *et*   *näe-ø*   *tema*   *ka*   *sa-i.*[www.epl.ee]
behind.LAT   that   look-IMP   s/he   also   get-PST.3SG
Lit. It would be especially nice to creep behind Friedman's back – hey, he got some too.
'It would be especially nice to hide behind Friedman because – hey, he got some too.'

(56) *se-da ma küll hea-ks ei kiida-ø ja*
this-PRT I indeed good-TRL NEG approve-CONNEG and
teise-ks autor sa ole-ks pida-nud oma-ø
second-TRL author you be-COND have-PST.PTCP own-GEN
mehe-ga selle-st rääki-ma, ja kui ei kõlba-ø
man-COM this-ELA talk-SUP and if NEG befit-CONNEG
jäta maha-ø ja võta-ø uus aga mitte et, hakka-d
dump-IMP and take-IMP new but not that begin-2SG
niimodi **seljataga** ...
like that back.behind.LOC
'I do not approve of this and secondly, the author, you should have talked to your husband about this and if he isn't good enough, dump him and get a new one but not do things behind backs.' [naistekas.delfi.ee]

It should be noted that lexicalization is the most important criterion when deciding between the simple and the complex structure. That is, the status of the complex item is primarily determined by its ability to express a holistic abstract meaning. Thus, as much as it is possible, the parameters of grammaticalization, e.g. those that are considered to indicate actualization of reanalysis (see sections 2.5.3.2 and 4.3) were not considered here.

The hybrid forms include such instances as where the body part related phrases have characteristics of both – the simple and the complex structure. For instance, in (57) the body part phrase carries exactly the same lexicalized meaning as in example (55), but the fact that there is an adjectival modifier (*lai* 'broad') between the LM (*tema* 's/he') and the phrase, precludes its analysis as a complex unit (see section 2.5.3.3.). Thus, structurally, it behaves as a regular simplex postposition complemented by a noun phrase, which consists of the head noun and an adjectival modifier.

(57) *Siis peida-Ø en-d* **tema-Ø** **laia-Ø** **selja-Ø**
then conceal-IMP you-PRT s/he-GEN wide-GEN back-GEN
taha *ja sõima Mihkelsoni-Ø edasi.*
behind.LAT and abuse-IMP MIhkelson-PRT on
'Then hide yourself behind his broad back and keep on calling Mihkelson names.' [www.epl.ee]

In the following, we will observe the distribution of these structures among the studied phrases. Figure 4 depicts the distributions of the simple and complex structure as well as the hybrid forms.

	Selja taga (back+behind)	Käe all (hand+under)	Külje all (side+under)	Jalge all (feet+under)	Käe kõrval (hand+beside)	Kaela peal (neck+on)
⊟ Free	4973	268	419	1050	7	128
☑ Unit	5966	4054	2530	867	771	87
∎ Hybrid	19	79	8	1	2	1
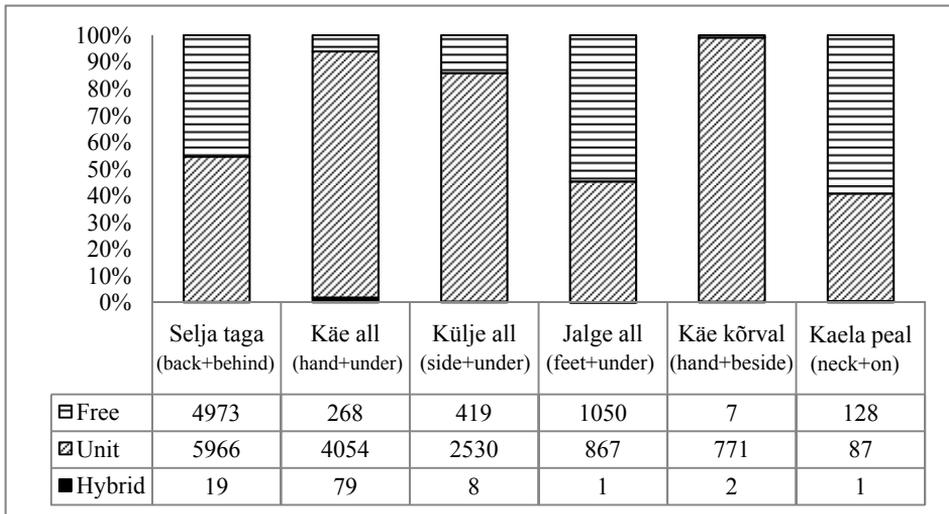
Figure 4. The distribution of the simple structures, the complex structures, and the hybrid forms among the studied phrases

It can be observed in Figure 4 that each of the studied phrases occurs in all three structures. The data suggest that in most cases the complex units were greater in number. It can be observed that *käe kõrval* (hand+beside), *käe all* (hand+under), and *külje all* (side+under) are used as complex units in 99% (771 examples out of 780), 92% (4054 examples out of 4401) and 86% (2530 examples out of 2957) respectively. However, two phrases are slightly more frequent as freely combined units. *Jalge all* (feet+under) and *kaela peal* (neck+on) occur as complex units in 45% (867 examples out of 1918) and 40% (87 examples out of 216) cases respectively. It is quite expected that as *kaela peal* (neck+on) as a less frequent phrase is also used less often as a complex unit because grammaticalization is usually associated with higher frequency (e.g. Hopper, Traugott (2003) [1993], Bybee 2010, 2007, 2003; Hoffmann 2005; Krug 2000). However, *käe kõrval* (hand+beside), which is also a less frequent item in this dataset, occurs as a complex unit in over 99% of the examples. Moreover, the most frequent of the studied phrases, *selja taga* (back+behind), occurs as a complex unit on 4983 occasions (54%). Thus, it seems that within the group of the studied phrases, there is no clear correlation between frequency and use as a complex unit.

However, perhaps, no clear correlation between high frequency and unit interpretation might have been expected for several reasons. First, that the development of complex function words observed here is an instance of gram-maticalization still in its very early stages. If high frequency is taken to be the result of grammaticalization, it is perhaps too early to expect a straightforward correlation at this point already. Second, it must be noted that the frequency of the phrases and the frequency of their use as a freely combined phrases and complex units is probably also dependent on each individual phrase and its

63

meanings. For instance, it can be assumed that the large proportion of usages as complex units amongst the phrases *käe all* (hand+under) and *käe kõrval* (hand+beside) are due to the fact that the contexts where these phrases may occur as free units are rather restricted, i.e. there are not so many entities that are described with a reference to one's hand. In contrast, the region to which *selja taga* (back+behind) refers is clearly much wider which may facilitate more frequent use of the phrase.

The data suggest that hybrid forms are generally infrequent – they form a larger group only among the data of *käe all* (hand+under), where they make up 2% (79 examples out of 4401) of the data. It seems that the frequency of such usages is also dependent on the semantics of the phrase. For instance, the body part term *käsi* 'hand' is frequently used with adjectival modifiers (*abistav käsi* 'helping hand', *kuldsed käed* lit. *golden hands* 'handy', *hoolas käsi* 'diligent hand') some of which also co-occur with postpositional phrases, e.g. *hoolsa käe all* (lit. under one's diligent hand).

4.3. Functions of the studied phrases

In this section, the functions of the studied phrases are discussed. It was stated above that based on the semantic and structural properties, each phrase may occur as the simple and as the complex structure and as the hybrid form. It was also mentioned in section 4.2 that when used as complex units, the phrases are more abstract than as the freely combined phrases. However, the phrases can be further divided based on the functions they carry. In the following, I will present the distribution of the functions for each phrase.

4.3.1. *Käe all* (hand+under)

The phrase *käe all* (hand+under) is used in three functions – BP+LOC (body part + simple locative postposition) which corresponds to the simple structure; MENTAL CONTROL and PHYSICAL CONTROL, which represent the usages as both, complex units and hybrid forms. The distribution of the functions of *käe all* (hand+under) are presented in Figure 5, which gives the absolute number as well as the percentages of instances that fall into each category.
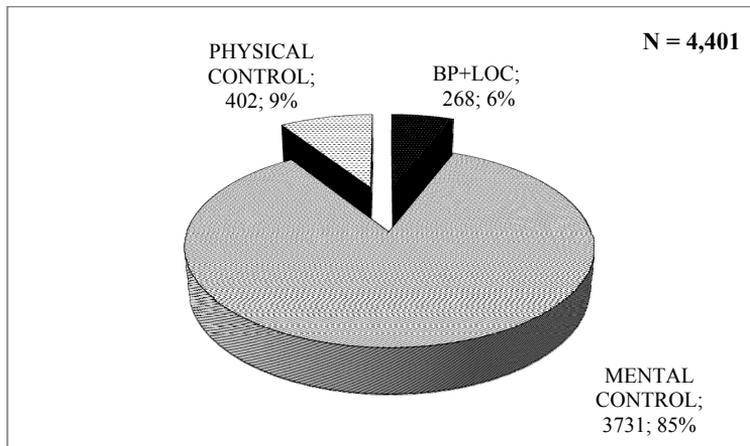
Figure 5. The distribution of the functions of *käe all* (hand+under) in the etTenTen corpus

It can be observed in Figure 5 that *käe all* (hand+under) was most often used to expresses MENTAL CONTROL. Examples of this function make up 85% of the data (3731 examples out of 4401). The two other categories – PHYSICAL CONTROL and BP+LOC – occur less frequently. Examples that belong to the former category make up 9% of the data (402 examples out of 4401) and examples that belong to the latter category make up 6% of the data (268 out of 4401). As in Estonian the body part term *käsi* 'hand' is not productively used to refer to object parts, the combination of the body part term and the simple locative function word make up all of the usages that represent the simple structure. Example (58) illustrates a typical example of such usages.

(58) *Ta*   *koba-s*   *seina-l*   *ja*   *ta-Ø*   **käe-Ø**
      s/he   grabble-PST.3SG   wall-ADE   and   s/he-GEN   hand-GEN
      alla   *jä-i*   *lambi-Ø*   *lüliti.*
      under.LAT   stay-PST.3SG   lamp-GEN   switch
       'She grabbed the wall and felt a light switch under his/her hand.'
      [www.poogen.ee.]

The other two categories – MENTAL CONTROL and PHYSICAL CONTROL – represent the lexicalized usages of *käe all* (hand+under) and are mostly analyzed to be complex units. In such cases, *käe all* (hand+under) does not express neither the body part meaning nor the locative function word, but the whole phrase expresses a more abstract relation, i.e. control.

   In cases where *käe all* (hand+under) expresses mental control, it typically portrays a relationship whereby one participant provides guidance to another or others (as in example (59)). In such cases *käe all* (hand+under) may also express an authoritarian relationship (as in example (60)). However, the usages illustrated in (59) and (60) are not always distinguishable, hence, such usages are considered to belong to the same category (MENTAL CONTROL).

(59) *1991   läk-s   Moskva-sse   filmi-ø   õppi-ma,   Nana*
      1991   go-PST.3SG   Moscow-ILL   film-PRT   study-SUP   Nana
      *Džordžadze-Ø   **käe-Ø**   **alla***.
      Džordžadze-GEN   hand-GEN   under.LAT
      Lit. In 1991 s/he went to Moscow, to study film under Nana Džordžadze's hand
      'In 1991 s/he went to Moscow to study film under Nana Džordžadze.'
      [www.ekspress.ee]

(60) *Riigi-Ø   esiprokuröri-Ø   Jüri   Pihli-Ø   edutamine*
      state-GEN   prosecutor-GEN   Jüri   Pihl-GEN   promotion
      kantsleritooli-le   peida-b   enda-s   justiitsminister
      chancellor's chair-ALL   conceal-3SG   itself-INE   minister of justice
      Ken-Marti   Vaheri-Ø   plaani-ø   tuu-a   kriminaalpolitsei
      Ken-Marti   Vaher-GEN   plan-PRT   bring-INF   criminal police
      *oma-Ø   **käe- Ø**   **alla.***
      own-GEN   hand-GEN   under.LAT
      Lit. Promotion of the state prosecutor Jüri Pihl to the position of chancellor
      conceals the plan of the Minister of Justice Ken-Marti Vaher to bring the
      criminal police under his own hand.
      'Promoting the state prosecutor to a position of chancellor is to cover up the fact
      that the Minister of Justice Ken-Marti Vaher wants to get the criminal police
      under his control.' [www.vnl.ee]

Sentences where *käe all* (hand +under) expresses physical control, usually
express situations where one participant physically manipulates someone or
something. It may refer to something being repaired or bettered (as in (61)), or
even that something is in (temporary) possession of someone (as in (62)).
However, the latter are quite rare.

(61) *Tüdruku-te   **käe-Ø**   **all**   valmi-nud   etnilis-te-s*
      girl-PL.GEN   hand-GEN   under.LOC   mature-PST.PTCP   ethnic-PL-INE
      kollektsiooni-de-s   või-b   leid-a   rõiva-i-d   erineva-te-le
      collection-PL-INE   might-3SG   find-INF   clothes-PL-PRT   various-PL-ALL
      suurus-te-le   nii   mees-te-le,   nais-te-le   kui   las-te-le /.../
      size-PL-ALL   as   man-PL-ALL   woman-PL-ALL   as   kid-PL-ALL
      Lit. The ethnic collections that have matured under the girls' hands include
      outfits for men, women, and children.
      'The ethnic collections made by the girls include outfits for men, women, and
      children.' [www.sirp.ee]

(62) *Hea ole-ks,   kui   esimene läbipääseja   võta-ks   oma-Ø*
     good be-COND   if   first   person through take-COND   own-GEN
     käe-Ø   alla   kogu   reisigrupi-Ø   skanneri-st   läbi
     hand-GEN   under.LAT whole   group-GEN   scanner-ELA   through
     sõit-nud   kraami-Ø,   muidu   või-b   üht-teist   kaotsi
     ride-PST.PTCP stuff-GEN   otherwise might-3SG   something   lost
     minn-a   /…/
     get-INF
     Lit. It would be good, if the first person through the security check takes the
     whole group's stuff under his/her hand as it emerges from the scanner, otherwise
     something might get lost
     'It would be good if the first person through the security check takes everybody's
     stuff into his/her possession; otherwise something might get lost.'
     [karavanserai.bluemoon.ee]

Both – MENTAL CONTROL and PHYSICAL CONTROL – are clearly based
on the simple structure, i.e. BP+LOC. The usages where *käe all* (hand+under)
behaves as a complex unit are interwoven. On the one hand, despite the fact that
(60) expresses MENTAL CONTROL, it bears a close meaning to (62) which
expresses PHYSICAL CONTROL, and can thus be interpreted as a kind of
abstract possession. On the other hand, (59) and (61) bear similar motivations as
both of them are based on the schema of forming/shaping something with one's
hands. This shift from BP+LOC to mental and physical control is harmonious
with general principles of grammaticalization and consistent with the findings
of other studies in Estonian as well as in other, non-related languages.

   Heine and Kuteva (2002: 166–167) report that in addition to various pos-
sessive markers[35], the body part term hand/arm or expressions that include the
body part term hand/arm have also developed into agent markers. For instance,
in Coptic the expression translated as 'on the hand' has developed into an agent
marker used in passive constructions (Heine, Kuteva 2002: 165). The Estonian
käe all (hand+under) cannot be analyzed as a passive marker but usages as
exemplified in (61) have a similar function as the examples that express
PHYSICAL CONTROL often occur in constructions where the agent (in this
case: *tüdrukud* 'girls') is somehow concealed. Coincidentally, the agent is also
the nominal modifier that precedes *käe all* (hand+under). Thus, in a way, *käe all*
(hand+under) introduces the agent in such examples. Moreover, there is
evidence from Mordvin, where the body part noun *ked'/käd'* is the source of
adposition that marks the performing agent of causative verbs and
POSSESSION as well as object transfer (van Pareren 2013: 100–101).

   Heine and Kuteva (2002: 167) have also reported Gardiner's (1957: 132)
findings from Egyptian where an expression meaning 'in my hand' has devel-
oped into a preposition meaning 'in the possession' and also 'in charge of', the
latter of which seems closest to the usages illustrated in (59) and especially (60).

---

[35]   For instance, in Bambara, Zande (reported by Kastenholz (1989), and Canon and Gore
[1931], (1952)), and Ewe.

Moreover, the Estonian simple function word *käes* that has developed from the body part term *käsi* 'hand' (< käe-s (hand-INE))[36] (Habicht 2000: 21), has two main functions in contemporary Estonian – possession and state (Ojutkangas 2001: 160). However, in case of animate landmarks, *käes* also expresses authority (see example (63)) (Ojutkangas 2001: 162). Such cases are considered to be functionally close to *käe all* (hand+under) when used to express MENTAL CONTROL. As will be demonstrated in section 4.5.1, the complex postpositional *käe all* (hand+under) also mostly occurs with animate landmarks. On the other hand, it is possible that the authoritative meaning is also affected by the other component of the source form, i.e. the simple postposition *all* 'under', which may induce the hierarchical relationship between the LM and the TR (see section 1.3). When compared to its simple counterpart *käes* (hand+INE), *käe all* (hand+under) as a complex item is definitely less grammatical. However, the motivation for its further grammaticalization still exists. The authoritative *käes* seems to have a negative connotation, but *käe all* (hand under) has a neutral or even positive connotation, as it mostly occurs in usages as exemplified in (59). Thus, it seems that there is a functional gap that *käe all* (hand+under) can fill.

(63)   *Peeter*   *vaevle-s*   *timuka-te*   ***käes***
       Peeter   pine-PST.3SG   executioner-PL.GEN   hand.LOC
       Lit. Peeter was struggling in the hand of the executioners.
       'Peeter was struggling due to the executioners.' [Ojutkangas 2001: 162; Rätsep 1978: 94]

4.3.2. *Külje all* (side+under)

As the body part term *külg* 'side' is also productively used as an object part, the free combinations of a noun and a simple postposition do not only include the category BP+LOC, but also combinations of object part nouns and simple locative postpositions (OP+LOC). In the instances where the phrase in interpreted as a unit (or very rarely, as a hybrid form), it expresses PHYSICAL PROXIMITY and MENTAL PROXIMITY. The distribution of the usage patterns is presented in Figure 6.

---

[36] The simple function word *käes* also comes in three-form sets (*kätte-käes-käest*) expressing the lative, locative, and separative.
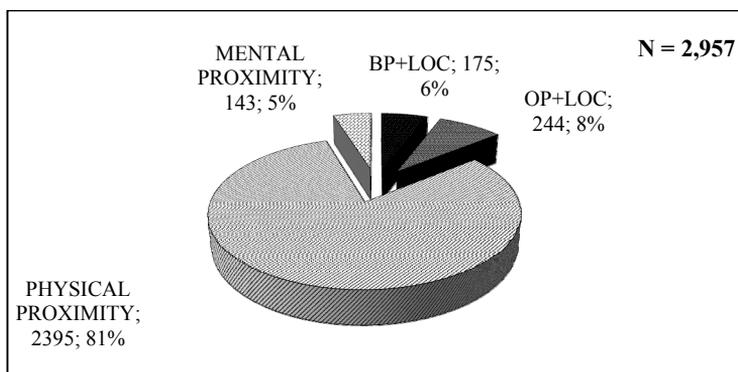
Figure 6. The distribution of the functions of *külje all* (side+under) in the etTenTen corpus

Külje all (side+under) is most often (81%; 2359 examples of 2957) used to express PHYSICAL PROXIMITY and less frequently in other functions. The examples that represent the function MENTAL PROXIMITY make up only 5% of the data (143 examples out of 2957); BP+LOC and OP+LOC make up 6% (175 examples out of 2957) and 8% (244 examples out of 2957) of the data respectively.

A typical example of the function BP+LOC is illustrated in example (64). Example (65) illustrates a usage where *külje all* (side+under) refers to a relation relative to an object.

(64) *Ma või-n selle-Ø teki-Ø ta-lle **külje-Ø alla***
     I   can-1SG this-GEN blanket-GEN s/he-ALL side-GEN under.LAT
     kinni pann-a ja ikka, tunnikese-Ø pärast on see tekk
     tight put-INF and still hour-GEN after be.3SG this blanket
     juba kaugele kaugele lüka-tud.
     already far far push-PST.IMPS.PTCP
     'I can put the blanket under her side, but still after a few hours the blanket has been pushed away.' [www.nupsu.ee]

(65) *Veel rohkem tekita-s hämmastus-t, et mööda sedasama-Ø*
     even more cause-PST.3SG surprise-PRT that along same-PRT
     *platvormi-ø sõit-is sinnasamasse lennuki-Ø **külje-Ø***
     platform-PRT drive-PST.3SG there aircraft-GEN side-GEN
     ***alla** vähemalt tosin erineva-t sõiduki-t.*
     under.LAT at least dozen different-PRT vehicle-PRT
     'It was even more suprizing that the same platform was used by a dozen vehicles to drive to just under the aircraft's side.' [algernon.ee]

As shown in section 4.2, the majority of the usages of *külje all* (side+under) represent the complex structure. In such cases, the phrase does not express the location of an entity relative to the side of another entity but rather the

proximity of the two entities in question. This is indicated by the context in which the phrase occurs. In (64) *külje all* (side+under) makes reference to a human being and the body part term clearly refers to the body part under which the blanket is tucked. In (65), the phrase is still interpretable as a freely combined unit – the the body part term refers to a part of an object (the aircraft), and the postpostion *alla* ('under') is also interpretable as expressing its literal meaning as the vehicles would probably be located on a lower level relative to the aircraft. In (66), however, the phrase appears in another context where it is used to express the relationship between a town and mountainbike trail. In this case, *külg* cannot be analyzed as a lexical item because the Estonian *linn* 'town' would not normally be divided into parts, which would be referred to with the word *külg* 'side'. In this case, the body/object part meaning has faded and the whole utterance expresses the notion of nearness. However, it is not impossible to interpret example (66) as also expressing a sense of proximity. Thus, the distinction between the two structures in some cases is to some extent intuitive (see also section 4.5.1.2).

(66) *5000   elaniku-ga   linna-Ø   **külje-Ø   alla**   jää-b*
      5000   dweller-COM   town-GEN   side-GEN   under.LAT   stay-3SG
      maailma-Ø   põneva-ma-i-d   mägirattarada-sid,
      world-GEN   exciting-COMP-PL-PRT   mountain bike trail-PL.PRT
      17   km   pikkune   Slickrock   Trail.
      17   km   length   Slickrock   Trail
      Lit. Under the side of the town of 5000 lies one of the most exciting mountain
      bike trails – the 17 km long Slickrock Trail.
      'Near the town of 5000 lies one of the most exciting mountain bike trails – the
      17 km long Slickrock Trail.' [reisiajakiri.gomaailm.ee]

However, *külje all* (side+under) does not only express PHYSICAL PROXIMITY, it can also express MENTAL PROXIMITY; the latter is illustrated in (67) where *külje all* is used to express being close to reality.

(67) *Võibolla   kunagi   edaspidi   kirjuta-b   keegi   ka*
      maybe   sometime   further   write-3SG   somebody   also
      artikle-i-d   kus   ilusa-te-st   lause-te-st   ja
      paper-PL-PRT   where   pretty-PL-ELA   sentence-PL-ELA   and
      käibetõde-de-st   jõu-ta-kse   reaalse-le   pisut   enam
      common knowledge-PL-ELA   reach-IMPS-PRS   reality-ALL   bit   more
      külje-Ø   alla.
      side-GEN   under.LAT
      Lit. Maybe sometime in the future, somebody will write a paper that reaches a
      bit more under the side of reality than the present one that preaches common
      knowledge with nicely formed sentences.
      'Maybe sometime in the future somebody will write a paper that is closer to the
      reality than the present one that preaches common knowledge in nicely formed
      sentences.' [sisekosmos.ee]

The shift from the body part related postpositional phrase to a complex unit that expresses PROXIMITY is rather expected and in line with the findings of other authors. Heine and Kuteva (2002: 139) have reported a similar development in ǁAni where the body part term *flank* has developed into a locative postposition meaning 'beside'. Svorou (1994: 72) reports a similar phenomenon in Abkhaz. Heine and Kuteva (2002: 271–272) list a cavalcade of similar instances with the term *side* from languages around the world, among them instances from Dullay (Amborn et al. 1980: 102), Bulu (Hagen 1914: 262), Kpelle (Westermann 1924: 12) and Tamil (Lehmann 1989: 122) where the body/object part that refers to *side* have developed into adpositions meaning 'next to' or 'near'. Although the Estonian simple function word *küljes* (side+INE)[37] typically expresses ATTACHMENT in a physical as well as an abstract sense, the Estonian Explanatory Dictionary also lists close proximity one of the meanings of the function word (68).

(68) *Liiklus   ol-i   väga   tihe –   auto   auto-ø   **küljes**   kinni.*[38]
      Traffic   be-PST.3SG   very   heavy   car   car-GEN attached   in
      Lit. The traffic was very heavy – a car stuck to the side of a car.
      'The traffic was very heavy – bumper-to-bumper.'

4.3.3. *Selja taga* (back+behind)

Selja taga (back+behind) is semantically one of the most interesting phrases in my data for several reasons. First, as it is the most frequent phrase with the strongest association between its components, it has more usage patterns than the rest of the studied phrases. Second, not all of the usage patterns are connected to each other. That is, not all of the patterns can be placed on a single diachronic continuum. However, all of the patterns are connected to the source form. Thus, it seems that in the complex unit *selja taga* (back+behind) has multiple branches of development, all of which start from the simple postpositional phrase result in a complex function word. *Selja taga* (back+behind) expresses six functions (see Figure 7).

---

[37] The simple function word *küljes* also comes in three-form sets (*külge-küljes-küljest*) expressing the lative, locative, and separative.
[38] http://www.eki.ee/dict/ekss/index.cgi?Q=k%C3%BCljes&F=M (Accessed 11.01.2016)
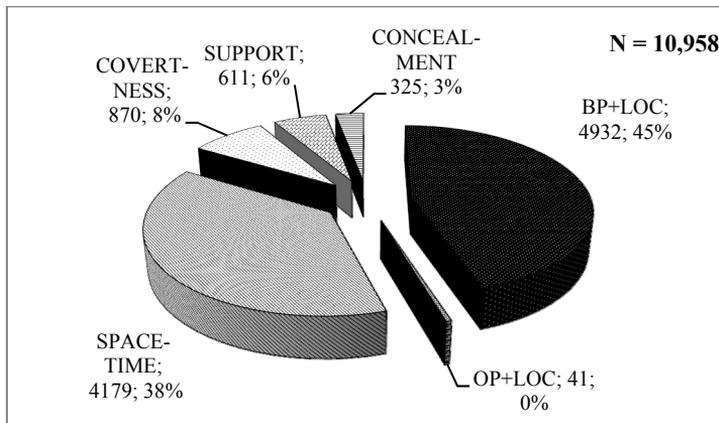
Figure 7. The distribution of the functions of *selja taga* (back+behind) in the etTenTen corpus

Selja taga (back+behind) is most commonly (45%; 4,932 out of 10,958) used in its literal meaning, i.e. BP+LOC (as exemplified in (69)). It is possible to use the body part term *selg* (back) to refer to object parts as well but such usages are not productive. The data suggest there were only 41 examples (less than 1%) where *selja taga* (back+behind) could be interpreted as referring to the posterior region of an object (as in (70)).

(69) *Näärivana   unusta-s   vitsakimbu-Ø   maha   kingikoti-Ø*
      Father Christmas   forget-PST.3SG   ferule-GEN   down   gift bag-GEN
      *aga   peit-is   tull-es   **selja-Ø**   **taha**.*
      but   hide-PST.3SG   come-GER   back-GEN   behind.LAT
      'Father Christmas forgot the ferule but hid the gift bag behind his back when
      coming.' [luuletus.www.ee]

(70) *Segesta   tempel   asu-b   üksildase-l   kallaku-l   ning*
      Segesta   temple   lie-3SG   lonely-ADE   slope-ADE   and
      *templi-Ø   **selja-Ø**   **taha**   jää-vad   ligipääsmatu-d*
      temple-GEN   back-GEN   behind.LAT   stay-3PL   invious-PL
      lubjakivimäe-d.
      lime stone mountain-PL
      Lit. The Segesta temple is situated on a lonely slope and behind the back of the
      temple remain invious lime stone mountains.
      'The Segesta temple is situated on a lonely slope and behind the temple there are
      invious lime stone mountains.' [www.novatours.ee]

However, in a little more than half of the examples, *selja taga* (back+behind) is lexicalized and expresses non-compositional meanings. The largest group among such usages is the category SPACE-TIME, which occurs on 4179 occasions (38%). This pattern comprises the uses where *selja taga* (back+behind) is used to express locative, temporal and ordinal functions. As these are not

always distinguishable, they are considered to fall under the same category and discussed together. The temporal function is exemplified in (71) where the phrase refers to a temporal relationship between the LM (Rooba) and the TR (nädal 'week'), which cannot be taken to be located behind the LM physically because it has a temporal dimension, i.e. it refers to a period of time. The locative function (exemplified in (72)) is very close to the compositional usage (BP+LOC) on the one hand, and to the temporal sense on the other hand. It is close to BP+LOC because they both express the posterior region, and it is hard to decide if the body part meaning has faded or not. However, the locative is distinguishable from the BP+LOC due to its similarity to temporal uses. When comparing the temporal *selja taga* in (71) and locative sense in (72), it can be seen that both of them use the same verb – *jääma/jätma* 'remain/leave'. Below, in section 4.6.3.1.3, it will be shown that *jääma/jätma* 'remain/leave' is particularly prone to co-occur with this function. The shift to temporal meaning is based on the metaphor TIME AS SPACE, more specifically, the Moving Observer metaphor because the past events (e.g. the week) are coded/projected as being left behind our backs (Lakoff, Johnson 1999; Veismann 2001). Similarly, there is an implied motion present in example (72) – the LM is projected as moving away from the TR (Soomaa), i.e. leaving it behind. No such motion is present in example (69) that illustrates the category BP+LOC.

(71)   *Tõepoolest,   juba   neli   aasta-t   põhjanaabri-te*
       Indeed   already   four   year-PRT   Northern neighbour-PL.GEN
       suurklubi-Ø   juures   treeni-nud   Rooba   mäng-is
       big club-GEN   at   train-PST.PTCP Rooba   play-PST.3SG
       üleeile   KalPa-Ø   vastu   suurepärase-Ø   mängu-Ø,
       day before yesterday KalPa-GEN   against   great-GEN   game-GEN
       *jät-tes   ühtlasi   **selja-Ø   taha**   eduka-Ø   nädala-Ø*
       leave-GER   also   back-GEN   behind.LAT successful-GEN   week-GEN.
       Lit. Indeed, Rooba who has played at a club of our Northern neighbours for four years now had a great game the day before yesterday, and also left behind a successful week.
       'Rooba who has been playing at a club of our Northern neighbours for four years now had a great game the day before yesterday, which also ended a successful week.' [www.epl.ee]

(72)   *Soomaa   jä-i   kiiresti   **selja-Ø   taha**,   sest*
       Soomaa   remain-PST.3SG   quickly   back-GEN   behind.LAT   because
       nüüd   on   tee-d   kiire-d   ja   laia-d.
       now   be.3SG   road-PL   quick-PL   and   broad-PL
       Lit. Soomaa stayed behind quickly because now the roads are quick and broad.
       'Soomaa was left behind quickly because now the roads are good and broad.' [www.aiaidee.ee]

In addition to locative and temporal usages, the category SPACE-TIME also includes a group of examples where *selja taga* (back+behind) expresses the

order of succession. In such cases *selja taga* (back+behind) usually refers to the ranking of individuals or teams in competitions (73), or the state of having surpassed something in a more abstract sense (74). They are considered to fall under the locative-spatial category because in this case *selja taga* (back+behind) is also prone to the verb *jääma/jätma* (74), however, it is not confined to it (73). Additionally, such usages are profiled through motion and therefore they are not always distinguishable from the locative and temporal *selja taga* (back+behind).

(73)   *Eestlas-te*   **selja-Ø**   **taha**   *mahtu-sid*   *esikümne-sse*
       Estonian-PL.GEN   back-GEN   behind.LAT   fit-PST.3PL   top ten-ILL
       veel   mitme-d   väga   tugeva-d   rattaorienteerumisriigi-d /…/
       more   many-PL   very   strong-PL   bike orienteering country-PL
       Lit. Behind the back of the Estonians, there were many good bike orienteering countries in the top ten
       'There were many more good bike orienteering countries in the top ten besides (after) Estonia.' [www.orienteerumine.ee]

(74)   *Toiduainetööstus*   *tõus-is*   *kolmanda-ks,*   *jät-tes*
       food industry   rise-PST.3SG   third-TRL   leave-GER
       selja-Ø   **taha**   *puidutööstuse-Ø*.
       back-GEN   behind.LAT   timberindustry-GEN
       Lit. The food industry rose to the third position leaving the timber industry behind its back.
       'The food industry surpassed the timber industry and rose to the third position.' [entsyklopeedia.ee]

The rest of the lexicalized usages of *selja taga* (back+behind) are divided between three relatively small and yet clearly distinct functional categories – COVERTNESS (8%, 870 occurrences), SUPPORT (6%, 611 occurrences), and CONCEALMENT (3%, 325 occurrences).

COVERTNESS includes examples where *selja taga* (back+behind) is used to refer to a deceitful or secretive manner in which something is done. The semantic shift is based on the inability of human beings to observe their posterior region (at all times). The events that occur behind a person's back can go unnoticed for that person. In examples of this category (as (75)), the locative interpretation has faded and the complex unit *selja taga* (back+behind) only expresses COVERTNESS.

(75)   *Et   Obama-Ø   administratsioon   on   tõepoolest   asu-nud*
       that   Obama-GEN   administration   be.3SG   indeed   set-PST.PTCP
       *oma-Ø   truu-de   liitlas-te   **selja-Ø**   **taga***
       one-GEN   loyal-PL.GEN   ally-PL.GEN   back-GEN   behind.LOC
       nen-de   arvel   sobingu-i-d   sõlmi-ma /…/
       they-GEN   expense   deal-PL-PRT   make-SUP
       Lit. The Obama administration has indeed started to make deals behind the back of their loyal allies.
       'The Obama administration has indeed started to make deals behind their loyal allies' backs.' [bhr.balanss.ee]

In the function SUPPORT, *selja taga* (back+behind) expresses the moral support towards the LM (as in (76)). The semantic shift is based on the physical conception of support – the person(s) that are located in one's posterior region may offer LM support in the physical sense as well as in more abstract ways, as can be observed in example (76). However, this pattern is mostly used to refer to purely moral support (as in (77)).

(76)　*… siis ütle-si-n　　ta-lle　　et　keegi　ei　austa-Ø*
　　　then　tell-PST-1SG　s/he-ALL　that　somebody　NEG　respect-CONNEG
　　　*sin-d　　sest　poisi-d　ol-i-d　　　kõik　mu-l　**selja-Ø***
　　　you-PRT　because　boy-PL　be-PST-3PL　all　I-ADE　back-GEN
　　　taga　　　*ja　kui　ta　ole-ks　　mu-lle　midagi　tei-nud*
　　　behind.LOC　and　if　s/he　be-COND　I-ALL　something　do-PST.PTCP
　　　ole-ksi-d　　　nad　ta-lle　　　tappa　　and-nud
　　　be-COND-3PL　they　s/he-ALL　beating　give-PST.PTCP
　　　Lit. Then I told him/her that no one respects them because all of the boys were standing right behind my back, and had s/he done anything, they would have given him/her a beating.
　　　'Then I told him/her that no one respects them because all of the boys were standing right behind me; had she done anything they would have given him/her a beating.' [www.lapsemure.ee]

(77)　*Ehk　　teis-te　　　sõna-de-ga:　　rahvas　on　　selgelt*
　　　that is　other-PL.GEN　word-PL-COM　people　be.3SG　clearly
　　　*Meikari-Ø　　**selja-Ø**　　**taga**　　ja　arva-b　　pigem,　et*
　　　Meikar-GEN　back-GEN　behind.LOC　and　think-3SG　rather　that
　　　justiitsminister　　valeta-b.
　　　minister of justice　lie-3SG
　　　Lit. In other words – people are clearly behind the back of Meikar and rather think that the Minister of Justice is lying.
　　　'In other words – people are clearly behind Meikar, and rather think that the Minister of Justice is lying.'
　　　[www.maaleht.ee]

The last function to be discussed here is CONCEALMENT, where *selja taga* (back+behind) is used to express state of the TR, which is described as being sheltered by the LM (as in (78)). The semantic shift is based on the conception of being physically protected or taken care of by somebody. However, as can be seen in example (79), this pattern also occurs in negative contexts where *selja taga* (back+behind) is used to refer to relation that the TR unrevealed by the LM (here: ACTA). Nevertheless, such examples are rare.

(78)　*Sellise-Ø　　mehe-Ø　　**selja-Ø**　　**taga**　　　on　　turvaline　ela-da.*
　　　This kind　man-GEN　back-GEN　behind.LOC　be.3SG　secure　live-INF
　　　Lit. It is secure to live behind the back of such a man
　　　'It is secure to live behind such a man.' [www.ekspress.ee]

(79) *Ajakirjaniku-d   või-ks   lõpuks   oll-a   nii   julge-d   ja*
      journalist-PL   can-COND   finally   be-INF   that   bold-PL   and
      rääki-da   viimaks   ka   se-da,   mis   on   ACTA-Ø
      speak-INF   at last   also   this-PRT what   be.3SG ACTA-GEN
      selja-Ø   taga   ja   tema-Ø   varjus.
      back-GEN   behind.LOC   and   s/he-GEN   shadow.LOC
      Lit. The journalists could finally come out and say what is behind the back of
      ACTA and in the lee of it.
      'The journalists should finally come out and say what is behind ACTA.' [viker-
      raadio.err.ee]

As stated above, all of the categories that include lexicalized uses of *selja taga* (back+behind) – spatio-temporal, COVERTNESS, SUPPORT, and CON-CEALMENT – do not seem to form a single chain of development. Instead, it seems that each one of them is the product of separate process of lexicalization. However, it is clear that the instances of *selja taga* (back+behind) that fall under the spatio-temporal category are related to each other. The shift from spatial to temporal domain is cognitively substantiated, and is one of the most basic functional shifts in grammaticalization (e.g. Heine, Kuteva 2002: 53). Nevertheless, it is hard to imagine the same type of relationship between the spatio-temporal category and other functional categories described. For instance, the category COVERTNESS is clearly connected to the pattern BP+LOC, but not to usages where *selja taga* (back+behind) is used to refer to temporal relations nor is it semantically related to any other patterns (SUPPORT or CONCEALMENT). Thus, it seems that the usages that fall under distinct categories are independent developments triggered by lexicalization of the usages that belong to BP+LOC.

However, the shift from BP+LOC to each functional category – SPACE-TIME, COVERTNESS, SUPPORT, and CONCEALMENT – is in line with the general principles of grammaticalization and similar developments can be found in other languages. For instance, the body part *back* is probably one of the most frequent sources of function words referring to the posterior region. Also, it is widely known that body part terms tend to develop into words that express spatial notions, which may then further develop to express temporal notions (c.f. Heine, Kuteva 2002: 47). Estonian *selja taga* (back+behind) has developed into a temporal marker that refers to events in the past relative to the Moving Observer (as in the case of the English *behind* in (80)). In addition, Svorou (1994: 158–159) describes a category termed POSTERIOR ORDER, which includes such temporal usages of posterior locative grams where the LM and TR are moving in the same direction and the TR comes after the LM (and also reaches the goal after the LM). The usages of *selja taga* (back+behind) exemplified in (73) and (74) can also be described in such terms.

(80)   *I'm so glad winter is finally **behind** us, but considering there was snow on the ground only 2 weeks ago, spring veggies aren't exactly on the table yet.*[39]

In addition to the spatio-temporal function, there is evidence of more abstract and also more specific functions that are associated with the body part *back*. For instance in Wolof, the body part term *gannaaw* 'back' expresses among other meanings the notion of *absence* (Moore 2000: 226). This notion is also expressed by utterances that include the body part *back* in English (as in example (81)). In this case, however, the utterance is not a grammatical item but rather an idiomatic expression. However, it expresses the same meaning as the instances of *selja taga* (back+behind) exemplified in (75). The rest of the functions – CONCEALMENT (82) and SUPPORT (83) – are also expressed by the English *behind*.

(81)   *… recently she has been telling me that our other friends have been talking about me **behind my back**, saying negative things.*[40]

(82)   *The singer, currently staying at his holiday home in Portugal, told friends he has no idea who is **behind** the 'completely false' accusation.*[41]

(83)   *I am 100% **behind** you. I know you'll find a way to achieve your dream* :)[42]

The Estonian body part term *selg* 'back' occurs in many phraseological expressions, but is not a source of any other function words besides *selja taga* (back+behind), except perhaps for *seljas* '(to have) on', which is used in connection with wearing clothing (see example (84)). However, it should be noted that the simple function word *taga* 'behind' has developed from a noun that refers to a hind part. However, the source form has disappeared from usage and the function word is no longer transparent. In contemporary language *taga* 'behind' is highly polysemous and also expresses also more abstract functions, such as SUPPORT (85), and CONCEALMENT (86), which are also expressed by *selja taga* (back+behind).

---

[39]   https://www.facebook.com/ChefNicoleGaffney/posts/805964169458153 (Accessed 11.01.2016)
[40]   https://psychologies.co.uk/my-friends-are-talking-about-me-behind-my-back (Accessed 11.01.2016)
[41]   http://www.dailymail.co.uk/news/article-2746719/I-no-idea-s-sex-claims-says-Sir-Cliff-Richard-insists-absolutely-hide.html (Accessed 11.01.2016)
[42]   http://www.baconismagic.ca/ecuador/the-secret-im-most-afraid-to-tell-you/ (Accessed 11.01.2016)

(84) *"Ilublogija   Liina   Ütt: teksa-sid   näe-b   minu-ø*
Beauty blogger   Liina   Ütt   jean-PL.PRT   see-3SG   I-GEN
***seljas   hea-l   juhu-l   paar   korda-ø   aasta-s!"*[43]
back.LOC   good-ADE   case-ADE   couple   time-PRT   year-INE
Lit. Beauty blogger Liina Ütt: you can see jeans on my back a couple of times a year at best!
'Beauty blogger Liina Ütt: you can see jeans on me a couple of times a year at best!'

(85) *Tead-is,   et   tema-ø   **taga**   on   rahvas.*[44]
know-PST.3SG   that   s/he-GEN   behind.LOC   be.3SG   people
Lit. S/he knew there were people behind him/her.
'S/he knew there were people supporting him/her.'

(86) *Teinekord   on   süütu-ø   lapsenäo-ø   **taga***
sometimes   be.3SG   innocent-GEN   child's face-GEN   behind.LOC
võimuahne   naine.[45]
power-hungry   woman
'Sometimes there is a power-hungry woman behind an innocent face of a child.'

The fact that the simple postposition *taga* 'behind' may carry a similar function as *selja taga* (back+behind) does not mean that *selja taga* (back+behind) is not a lexicalized holistic unit. For instance, in the examples (76)–(78) *selja taga* (back+behind) could not be substituted with the simple form *taga* 'behind' without altering the meaning. In example (79), the switch would be possible because there are enough contextual hints present. Thus, although *taga* 'behind' and *selja taga* (back+behind) to a certain extent carry the same function, they cannot be replaced by each other in all contexts. Additionally, *selja taga* (back+behind) is more complex, less grammaticalized and therefore more specific than *taga* 'behind', which is more polysemous, abstract, may require more context. A need for a more specific means of expression may account for the use of *selja taga* (back+behind) in these functions. However, not all of the functions of *selja taga* (back+behind) can be expressed by *taga* 'behind' alone. For instance, *taga* 'behind' cannot be used in the temporal and locative functions exemplified in (71), (72) or in (74). A similar train of thought has been put forward by Sigurd (1993), who also claims that … "multi-word prepositions are often specifications of simple prepositions" (1993: 204). The more specific meaning is achieved by adding the noun (ibid.).

---

[43]   http://www.ohtuleht.ee/674765/ilublogija-liina-utt-teksasid-naeb-minu-seljas-heal-juhul-paar-korda-aastas (Accessed 11.01.2016)
[44]   http://www.eki.ee/dict/ekss/index.cgi?Q=taga&F=M (Accessed 11.01.2016)
[45]   http://www.eki.ee/dict/ekss/index.cgi?Q=taga&F=M (Accessed 11.01.2016)

4.3.4. *Käe kõrval* (hand+beside)

Käe kõrval (hand+beside) is used to express three functions – BP+LOC, BESIDE and ACCOMPANIMENT. The distribution of the functions is presented in Figure 8.
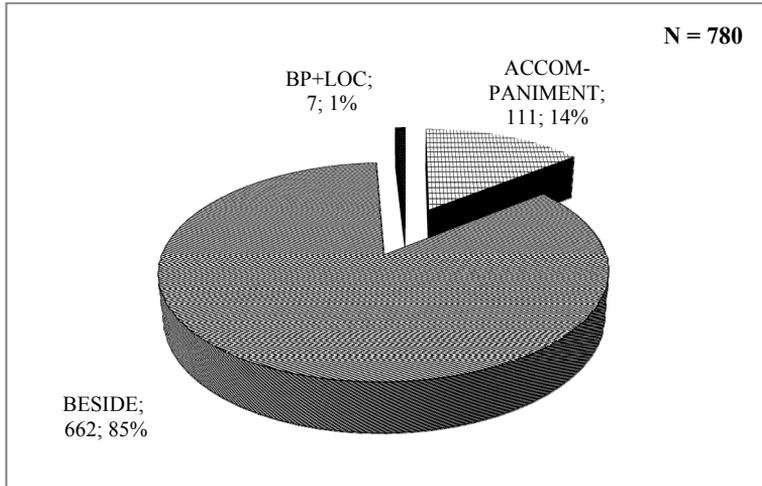
Figure 8. The Distribution of the functions of *käe kõrval* (hand+beside) in the etTenTen corpus

Käe kõrval (hand+beside) barely occurs in its compositional meaning. Such usages (as in (87)) make up barely 1% of the examples (7 instances). This is quite expected given the specificity of the compositional meaning (beside one's hand) and the fact that the Estonian explanatory dictionary lists it as a complex unit that functions as an adverb and a postposition.[46]

(87)　*Mu-Ø*　　*parema-Ø*　　**käe-Ø**　　　**kõrvale**　　　*ilmu-s*
　　　　I-GEN　　right-GEN　　hand-GEN　　beside.LAT　　appear-PST.3SG
　　　　jääkülm　　*kristallklaas*　　*vee-ga.*
　　　　ice cold　　crystal glass　　water-COM
　　　　'Next to my right hand appeared a crystal glass with water.'

The instances of *käe kõrval* as complex units are divided into two inter-related functions – BESIDE and ACCOMPANIMENT. BESIDE, which occurs in 662 instances (85%), is the most frequent function. This category comprises such cases as where *käe kõrval* (hand+beside) is used to describe the location of an animate or inanimate TR in the side region of the LM (as in examples (88) and (89)). The lexicalized meaning of *käe kõrval* (hand+beside) is not particularly far from the compositional meaning. However, when *käe kõrval* (hand+beside)

---

carries the function BESIDE, the body part term meaning is less prevalent, and the whole phrase carries a locative function. Moreover, in this pattern the LM and TR are typically attached to each other, which is not necessarily the case with the freely combined phrases.

(88) *Väike   Lissi   astu-s   arglikult   ema-Ø   **käekõrval***
Little   Lissi   step-PST.3SG   shyly   mother-GEN   beside.LOC
koolimaja-Ø   peaukse-st   sisse.
school house-GEN   front door-ELA   inside
Lit. Little Lissi shyly stepped in through the front door of the school house beside her mother's hand.
'Little Lissi shyly stepped in through the front door of the school house beside her mother.' [blablabla.ee]

(89) *Enamus   maa-d   lükka-si-n   veel   jalgratas-t   **käekõrval**.*
Most   way-PRT   push-PST-1SG   also   bicycle-PRT   beside.LOC
Lit. Most of the way I pushed the bicycle beside my hand.
'Most of the way I pushed the bicycle beside me.' [www.geopeitus.ee]

However, sometimes *käe kõrval* (hand+beside) is used to refer to contexts where attachment of the LM and the TR is improbable, or even inappropriate. Such context is exemplified in (90) where the attachment is unlikely and the literal meaning does not make sense. Rather, the phrase is used to mean that s/he took the owner of the company *with* them. In such cases, the locational meaning is fading and the sense of accompaniment (91) is setting in.

(90) *Alles   siis,   kui   ma   firma-Ø   ühe-Ø   omaniku-Ø   Priit*
Only   then   when   I   company-GEN   one-GEN   owner-GEN   Priit
*Alamäe-Ø   **käekõrvale**   võt-si-n   ja   tema-ga   mööda*
Alamäe-GEN   beside.LAT   take-PST-1SG   and   s/he-COM   along
tulevas-t   linnaku-t   ringi   sõit-si-n /…/
future-PRT   site-PRT   around   drive-PST-1SG
Lit. Only when I took one of the owners beside my hand and we drove along the site …
'Only when I took one of the owners with me to drive around the site.'
[www.director.ee]

(91) *Kierkegaardi-Ø   **käekõrval**   tee-n   ma   siin   katse-t*
Kieregaard-GEN   hand.beside.LOC   make-1SG   I   here   try-PRT
minn-a   teis-t   tee-d.
go-INF   another-PRT   way-PRT
Lit. Beside the hand of Kierkegaard, I am trying to go another way.
'(Together) with Kierkegaard I am trying to go another way.' [www.eestikirik.ee]

Based on the lexical component *käsi* 'hand', it is difficult to find a parallel in other languages to the semantic shift that *käe kõrval* (hand+beside) has gone

through. In this case, it seems that the simple postposition *kõrval* (ear+ADE)[47] plays a significant role in the development of the holistic meaning of the phrase. *Kõrval* can be used to express the locative BESIDE as well as ADDITION. However, the body part term *käsi* 'hand' has also a role to play in this. BESIDE as well as ACCOMPANIMENT also include a sense of GUIDANCE meaning that these functions resemble that of *käe all* (hand+under) discussed in section 4.3.1 above. Thus, BESIDE and ACCOMPANIMENT do not express the nearness or accompaniment of a random person but someone who would guide and lead the TR.

4.3.5. *Kaela peal* (neck+on)

Kaela peal (neck+on) occurs in three functions – OP+LOC, BP+LOC and BURDEN. The distribution of the functions of *kaela peal* (hand + beside) is presented in Figure 9.
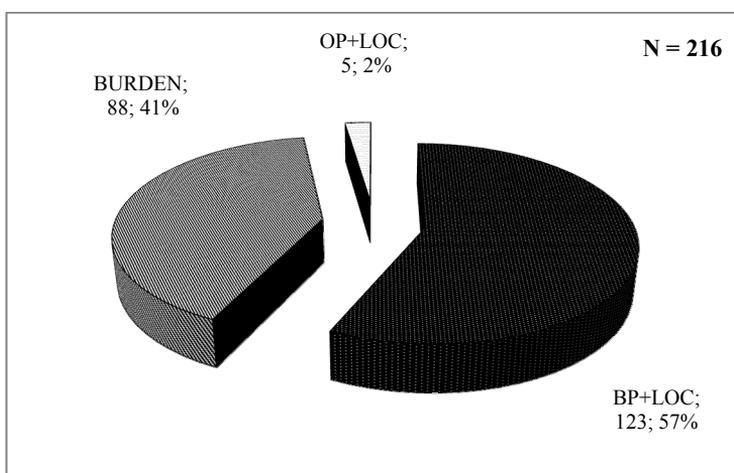
Figure 9. The distribution of the functions of *kaela peal* (neck+on) in the etTenTen corpus

Kaela peal (neck+on) is most frequently (57%, 123 examples out of 216) used as a free combination of a body part noun and simple postposition (as in example (92)). OP+LOC (example (93)), which is also taken to represent the simple structure, occurs very infrequently in the data (2%; 5 examples out of 216).

---

[47] The simple function word *kõrval* also comes in three-form sets (*kõrvale-kõrval-kõrvalt*) expressing the lative, locative, and separative.

(92) *Mure   selle-s,   et   mu-Ø   1   kuuse-l   neiu-l*
problem   this-INE   that   I-GEN   1   month old-ADE   girl-ADE
on   mingi-d   imeliku-d   täpikese-d   näo-s   ja
be.3SG   some-PL   weird-PL   spot-PL   face-INE   and
*mõni   üksik   ka   **kaela-Ø**   **peal**.*
some   single   also   neck-GEN   on.LOC
'I have a problem – my one month old daughter has weird spots on her face and
a few on her neck.' [www.nupsu.ee]

(93) *Jätka-Ø   **kaela-Ø**   **peal**   liikumis-t   ülespoole   kuni*
continue-IMP   neck-GEN   on.LOC   moving-PRT   upwards   until
14. astmetraadi-ni   ja   tule-Ø   samamoodi   tagasi.
14th fret-TER   and   come-IMP   likewise   back
'Continue moving on the neck until the 14th fret, and come back the same way.'
[www.kitarr.ee]

As a complex unit *kaela peal* (neck+on) carries the function BURDEN.
Although BURDEN is semantically rather specific – it is an abstract meaning
that is expressed by the phrase as a whole. BURDEN is illustrated in example
(94) where *kaela peal* expresses a more abstract relationship between the LM
(abikaasa 'spouse') and the TRs (*maamaja* 'country house', *laps* 'child' and
ema 'mother').

(94) *Abikaasa-l   on   **kaela-Ø**   **peal**   maamaja,   väike   laps*
spouse-ADE   be.3SG   neck-GEN   on.LOC   country house   small   child
ja   haige   ema   /…/
and   sick   mother
Lit. My spouse has a country house, a small child, and a sick mother on his/her neck
'My spouse has a country house, a small child, and a sick mother around her
neck.' [www.virumalev.ee]

It is clear that such usages as exemplified in (92) do not provide suitable con-
texts for the development of usages as exemplified in (94). Rather, the moti-
vation for the semantic shift is reflected in usages as exemplified in (95), where
the comparative construction *veskikivina* 'as a millstone' refers to the connec-
tion between the abstract burden and the physical burden. Thus, it seems that
the function BURDEN is connected to BP+LOC. However, the example (95)
remains the only one in the data that includes physically heavy objects around
someone's neck. Typical examples of the pattern BP+LOC describe smaller,
light objects on one's neck (often medical problems) (see example (92)). This
may be owing to the nature of the texts – a considerable amount of data of *kaela
peal* (neck+on) originates from forums where health problems are discussed.
However, given that the contexts that may give rise to the development of the
complex unit (as 95) are quite rare, based on synchronic analysis, the motivation
for the development remains uncertain. This question will be revisited in
diachronic analysis of *kaela peal* (neck+on).

(95) *Kui   esikesik   Edik   ministri-na   raudtee-Ø*
      when   first centralist   Edik   minister-ESS   railway-GEN
      taasriigista-s,   tead-is   ta   hästi,   et   see
      re-nationalise-PST.3SG   know-PST.3SG   s/he   well   that   it
      *tuleviku-s   Eesti-le   veskikivi-na   **kaela-Ø   peale**   jää-b.*
      future-INE   Estonia-ALL   millstone-ESS   neck-GEN   on.LAT   stay-3SG
      Lit. When the Central Party leader Edik re-nationalized the railway as a minister,
      he knew that it would be a millstone on the neck of Estonia.
      'When the Central Party leader Edik re-nationalized the railway as a minister, he
      knew that it would become a millstone around Estonia's neck.'
      [majandus.delfi.ee]

Because of its high specificity, it is difficult to find parallel developments to
that of *kaela peal* (neck+on) from other languages. However, there are similar
semantic shifts that concern lexical items, for instance, the English expression *a
millstone around one's neck*. However, there is a similar development in Esto-
nian whereby the body part term *kael* 'neck' has developed into the simple
function word *kaelas* (neck+INE)[48] that carries the same function (96). It will be
demonstrated in section 4.8.5.1 that because the simple and the complex expres-
sion with the body part term *kael* 'neck' carry a similar function, they might
have influenced each other's development.

(96) *Iga-l   viiteistkümnenda-l   eestlase-l   on   pikaajaline*
      every-ADE   fifteenth-ADE   Estonian-ADE   be.3SG   long-term
      *võlg   **kaelas**.*[49]
      debt   on.LOC
      Lit. Every fifteenth Estonian has a long-term loan on their neck.
      'Every fifteenth Estonian has a long-term loan on them.'

4.3.6. *Jalge all* (feet+under)

Jalge all (feet+under) carries three functions – BP+LOC, OPPRESSION and
INCEPTIVENESS. The distribution of the functions of *jalge all* (feet+under) is
presented in Figure 10.

---

[48]   *Kaelas* (neck+INE) also has three locative forms (*kaela-kaelas-kaelast*) expressing the
lative, locative, and separative.
[49]   http://www.postimees.ee/2952437/igal-viiteistkumnendal-eestlasel-on-pikaajaline-volg-
kaelas (Accessed 11.01.2016)
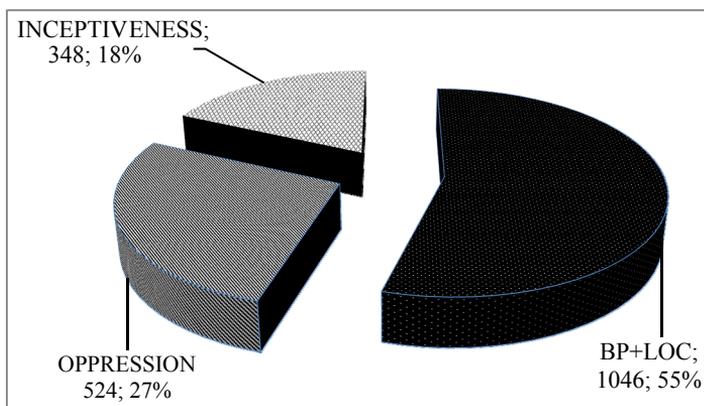
Figure 10. The Distribution of the functions of *jalge all* (feet+under) in the etTenTen corpus

Figure 10 shows that *jalge all* (feet+under) is most frequently (55%, 1046 occasions out of 1918) used in its literal meaning, i.e. the pattern BP+LOC (as in (97)). However, *jalge all* (feet+under) also occurs as a lexicalized item. In this case, it expresses OPPRESSION (27%, 524 out of 1918) (as in (98)) or INCEPTIVENESS (18%, 348 out of 1918) (as in (99)).

(97) *Ta   peatu-s   siis   ja   tund-is   mulda-ø*
      s/he   stop-PST.3SG   then   and   feel-PST.3SG   soil-PRT
      ***jalg-e   all**.*
      foot-PL.GEN   under.LOC
      'S/he stopped and felt the soil beneath his/her feet.' [www.epl.ee]

(98) *Majandus   jät-ku   talla-ma-ta   **jalgealla**   keskkond,*
      economy   leave-JUS   tread-SUP-ABE   feet.under.LAT   environment
      mille-s   hakka-vad   ela-ma   meie-Ø   lapse-d.
      what-INE   start-3PL   live-SUP   we-GEN   kid-PL
      Lit. The economy should not tread under feet the environment that our kids will
      live in.
      'The economy should not tread upon the environment in which our kids will
      live.' [kampaania.erakond.ee]

(99) *haiguse-d   hakka-vad   ne-i-d   kimbuta-ma   alles   siis   kui*
      disease-PL   start-3PL   they-PL-PRT   bug-SUP   yet   then   when
      *kohtutee   **jalgealla**   satu-b.*
      court way   feet.under.LAT   happen-3SG
      Lit. Diseases will not bug them before the court way happens up under feet
      'They will not get sick before they have to embark upon their way to court.'
      [www.maaleht.ee]

Similarly to *selja taga* (back+behind), the developmental paths of OPPRESSION and INCEPTIVENESS are not connected. It is difficult to find corresponding developments in Estonian or other languages because in this case we are dealing with highly specific functions. It will be demonstrated below (sections 4.6.6.1 and 4.6.6.2) that the usages exemplified in (98) and (99) rather dependent on the immediate sentential context, and in this respect *jalge all* (feet+under) is definitely the least grammaticalized of the studied phrases.

4.3.7. Summary and main conclusions
of the functions of the studied phrases

The data suggest that all of the studied phrases are analyzable as linguistic elements, which in addition to their literal interpretations, also carry meanings that are not derivable from the meanings of their components. In such cases, the phrases carry holistic meanings and are interpreted as more abstract than the literal interpretations. Such usages include examples where the semantics of the lexical component (the body part term) has faded and the phrase as a whole is used to expresses an abstract relation.

In some cases, these abstract meanings are rather expected as they converge with the developments of similar source forms in other languages. For instance, the development of *selja taga* into a complex unit that expresses the functions SPACE, TIME, and POSTERIOR ORDER is rather expected, as is the development of *käe all* (hand+under) into a function word that expresses MENTAL OR PHYSICAL CONTROL, or *külje all* (side+under) into a complex unit that expresses PROXIMITY. These shifts are in line with the general principles of grammaticalization and have been documented in many languages. In some cases, however, the complex units express rather specific meanings. For instance, *kaela peal* (neck+on) as a complex unit is used to express the notion BURDEN, which is clearly a more abstract concept than the combination of the body part term and a simple locative gram, but is also quite specific. Thus, it is difficult to find parallels to such a development in other languages because currently *kaela peal* (neck +on) cannot be associated with a broader category. The same applies for *jalge all* (feet+under), which expresses quite specific notions as a complex unit. As expected, greater specificity mostly affects the less frequent phrases. This was especially the case for *jalge all* (feet+under) and *kaela peal* (neck+on). It will be shown below that *jalge all* (feet+under) shows lesser degree of grammaticalization almost in every other aspect too.

However, it seems that high specificity may be regarded as a characteristic feature of Estonian postpositions in general (see also section 2.2). This is probably connected to the fact that the development of Estonian function words involves an intermediate stage of lexicalization (see section 2.3), and that the adverbial and postpositional uses of Estonian function words may be rather close (see section 2.2). This also applies to complex function words and is probably even more relevant among complex units, because of their structural

complexity and relatively recent development. Complex units include more lexical and grammatical content because they are currently analyzable on many levels. A complex unit includes the body part term and a simple postposition, which in some cases (*kõrva-l* (ear-ADE), *pea-l* (head-ADE)) are still transparent, yet have their own grammatical meaning ('beside', 'on'). Although it was argued that as complex units, the phrases express holistic meanings, it probably the case that these holistic meanings are affected by the components of the phrase, in the sense that as long as such units are transparent and morphologically complex (on many levels), it could not be expected that they can be associated with very broad abstract functions.

4.4. The complex structure –
complex adverbs and complex postpositions

In this section, I take a closer look at the examples that are analyzed as complex units. As mentioned in section 4.2, as complex units the phrases may occur as complex adverbs or complex postpositions. The distribution of these parts-of-speech among the studied phrases is given in Figure 11.
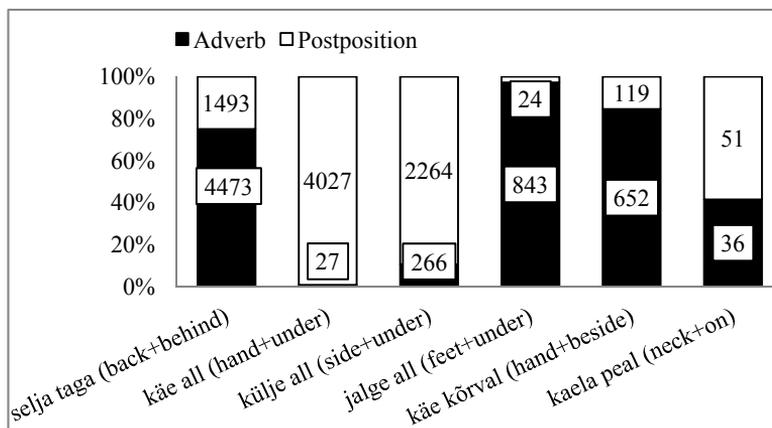
Figure 11. The distribution of the complex adverbs and complex postpositions among the studied phrases in the etTenTen corpus (N = 270,000,000)

The data suggest that as complex items, two of the studied phrases are preferred to be used as adverbs and three of the phrases seem to be inclined towards being used as postpositions. It can be observed in Figure 11 that *käe all* (hand+under) and *külje all* (side+under) occur almost without an exception as postpositions. In the case of these phrases, the adverbs make up less than 1% (27 examples out of 4054) and 10% (266 examples out of 2530) of the examples respectively. The less frequent *käe kõrval* (hand+beside) and intermediately frequent *jalge all* (feet+under) are both clearly preferred to occur as adverbs when used as complex units. The amount of postpositional uses remains around 15% (119

examples out of 771) in the case of *käe kõrval* (hand+beside) and 3% (24 examples out of 867) in the case of *jalge all* (feet+under). 75% (4473 examples out of 5966) of the complex units of the most frequent phrase *selja taga* (back+behind) are used as adverbs. Nevertheless, the distribution of adverbial and postpositional uses is more even in the case the of *kaela peal* (neck+on) with postpositional uses a little more frequent at 59% (51 examples out of 87).

As discussed in sections 4.3.3 and 4.3.6, both *selja taga* (back+behind) and *jalge all* (feet+under) carry functions that do not share a common developmental path. Thus, in the following, I discuss the distribution of adverbs and postpositions among the functions of *selja taga* (back+behind) (see Figure 12) and *jalge all* (feet+under) (see Figure 13) separately.
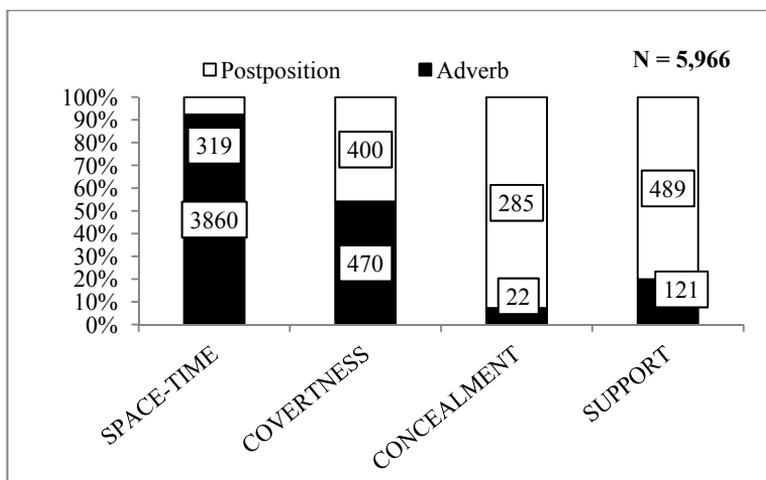
Figure 12. The distribution of the complex adverbs and complex postpositions among the different functions of *selja taga* (back+behind) in the etTenTen corpus (N = 270,000,000)

Figure 12 shows that the distribution of adverbial and postpositional use varies greatly between the different functions of *selja taga* (back+behind). For instance, the spatio-temporal *selja taga* (back+behind) is predominantly (92%; 3860 examples out of 4179) used as an adverb. That means that only 8% of the examples are used as postpositions. Although I will henceforth not distinguish the usages that make up the spatio-temporal function (see 4.3.3.), it should be noted that the postpositional uses mostly occur within examples where *selja taga* (back+behind) refers to an order (as in example (73) above). The data show that COVERTNESS is the only category where *selja taga* (back+behind) occurs equally as frequently as a postposition (46%; 400 examples out of 870) and an adverb (54%; 470 examples out of 870). The rest of the categories – CONCEALMENT and SUPPORT – are inclined to be used as postpositions. CONCEALMENT occurs as a complex postposition in 92% (285 examples out of 307) of cases, and SUPPORT occurs as a complex postpositional in 80% of

cases (489 examples out of 610). It seems that CONCEALMENT and SUPPORT strongly prefer the postpositional use because the postpositional construction allows expressing the LM more explicitly (as a preceding (pro)noun). In these functions, the LM expresses the concealer and the supporter, which are taken to be important participants in this relationship. It is possible that in the case of COVERTNESS and SPACE/TIME, the LM is more readily also deductible from the context.
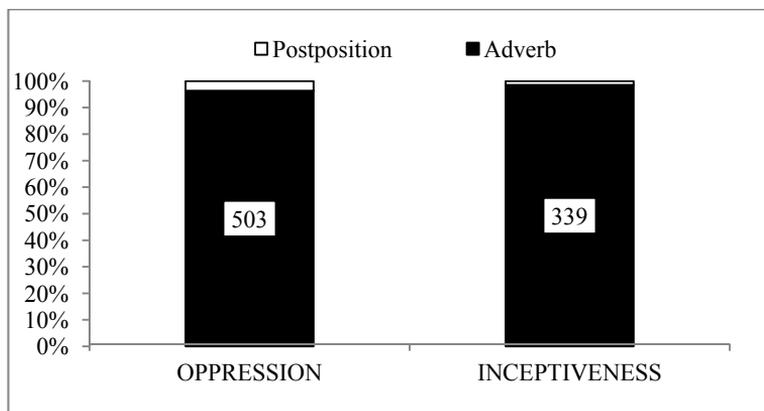
Figure 13. The distribution of the complex adverbs and complex postpositions among the different functions of *jalge all* (feet+under) in the etTenTen corpus (N = 270,000,000)

Figure 13, which depicts the distribution of complex adverbs and complex postpositions among the examples of *jalge all* (feet+under), shows that both categories – OPPRESSION and INCEPTIVENESS – clearly prefer to occur as complex adverbs. Adverbial uses make up 96% (503 examples out of 522) and 98% (339 examples out of 344) among OPPRESSION and INCEPTIVENESS respectively. In both cases, the postpositional uses are extremely rare.

4.5. Reanalysis of the studied phrases

The reanalysis of the studied phrases is observable in two features – the semantic class and the grammatical number – of the (pro)noun that precedes the body part related phrase (see section 2.5.3.2). Thus, in this chapter, I only analyze such usages where the phrase is preceded by a noun or a pronoun that belongs to the same phrasal structure as the phrase. In the case of freely combined units, the (pro)noun behaves as a modifier of the body part term; in the case of the complex postpositions, the preceding (pro)noun (PN) behaves as a complement of the complex postposition (see sections 2.5.3.2 and 1.3).

Because the following analysis is based on the feature of the preceding (pro)noun, I will briefly introduce the principles that were established when coding the relevant data:

i   The grammatical number is coded as a formal factor:
    –   In cases where the phrase was preceded by two coordinated (pro)nouns – *kunstniku ja käsitöölise käe all* (artist.GEN and craftsman.GEN hand +under) – the lemma was coded as the lemma closest to the phrase (*käsitööline* 'handicraftsman'). The reason being that such cases may be interpreted as examples of elliptical coordination – *kunstniku käe all ja käsitöölise käe all* 'under the artist's hand and under the craftsman's hand' – where the first instance of *käe all* (hand+under) is omitted. The omission is possible in case of both – the simple and the complex structure. This is also the reason why such cases are not considered to indicate semantic incompatibility with the simple structure.
    –   In cases where the PN is itself preceded by a numeral – e.g. *67 kunstniku käe all* (67 artist-GEN.SG hand under; lit. 'under the hand of 67 artists') – were coded as singular forms. Although such usages refer to plurality in the extra-linguistic reality, the quantifier phrase requires the noun to be in the singular form. As using the plural form here is not possible under any circumstances, such cases are not considered to suggest semantic incompatibility and are coded based on the formal marker. Following the same principle, the collective PNs (e.g. *meeskond* 'team') are coded as singular forms. Such usages are considered to indicate semantic incompatibility with the simple structure, but this is reflected by the semantic class of the PN and coded as such accordingly.

ii   The semantic class is coded as a semantic factor:
    –   In cases where the PN is expressed with a pronoun, the semantic class is coded based on the semantic class to which that pronoun refers. For instance, in the utterance *meeskond, kelle käe all* (team who.GEN hand +under; lit. under the team's hand), the PN lemma is coded as *kes* 'who' and the semantic class as collective, because the pronoun refers to the team.

4.5.1. The semantic class of the preceding (pro)noun

In this section, the semantic class of the lemmas of the preceding noun (PN) are discussed. As suggested in section 2.5.3.2, the studied phrases occur in contexts that would not be semantically compatible in cases of the simple structure, and hence they are considered to indicate that the simple structure has been re-analyzed as a complex structure.

In the following section, it will be observed:

i   which semantic classes are represented by the PNs of each phrase;
ii   what is the distribution of semantic classes among the simple and the complex structures and among the individual functions (in the case of *selja taga* (back+behind) and *jalge all* (feet+under));
iii how large is the proportion of uses that indicate contextual expansion;
iv   what does the distribution of semantic classes reveal about the developmental paths of the studied phrases?

4.5.1.1. *Käe all* (hand+under)

In this section, analysis of the semantic class of the PN lemma of *käe all* (hand+under) is presented. The semantic classes that are present among the PNs in the examples of *käe all* (hand+under) are presented in Table 4.

Table 4. The distribution of the semantic classes of the PN among the examples where *käe all* (hand+under) occurs as a freely combined phrase, a complex unit, and a hybrid form

Semantic class of the PN	Total	Free	Unit	Hybrid
Human	4104 (98%)	86 (99%)	3964 (98%)	54 (68%)
God	25 (1%)	1 (1%)	5 (<1%)	19 (24%)
Collective	64 (2%)	0	58 (1%)	6 (8%)
Total	**4193 (100%)**	**87 (100%)**	**4027 (100%)**	**79 (100%)**

The phrase *käe all* (hand+under) occurs with three semantic classes – human, God, and collective. Quite expectedly, the human PNs are the most common semantic class, such examples make up most (98%) of the data. In the case of *käe all* (hand+under), the PNs that refer to God are presented as a separate semantic class that makes up 1% of the data (25 examples). Although the Christian God is usually portrayed in the form of a human being, the distinction between PNs that refer to human beings and God is not necessary from the point of view of the expansion. However, as the examples where the PN refers to God form a distinct pattern, which has its own characteristics, they are considered to form a distinct class. The third class – collective PNs – mostly includes human collectives (100)) and institutions (101), but also few abstract entities (102) that were too rare to form a distinct group. However, collective PNs are not frequent either. Table 4 shows that the examples where the PN refers to a collective make up only 2% (64 examples) of all the uses of *käe all* (hand+under). Thus, it can be observed that *käe all* (hand+under) is mostly used with human PNs. In the following section, the distribution of semantic classes is viewed among the freely combined phrases, the complex units, and the examples that have characteristics of both structures.

(100) *Kuidas   see   **Kukke-de   dünastia-ø   käe-ø   alla***
how   this   Kukk-PL.GEN   dynasty-GEN   hand-GEN   under.LAT
sattu-des   teisiti   oll-a   saa-ks-ki.
happen-GER   otherwise   be-INF   get-COND-CL
Lit. How else could it be having come under the hand of the Kukk dynasty.
'How else could it be having come under the control of the Kukk dynasty.'
[www.horsemarket.ee]

(101) *Chrysler,   kes   on   nüüdse-ks   juba   üle   aasta-ø*
Chrysler   who   be.3SG   current-TRL   already   over   year-GEN
*Itaalia-ø   autotööstuse-ø   suurnime-ø   **Fiat   SpA-ø***
Italy-GEN   automotive industry-GEN   big name-GEN   Fiat   SpA-GEN
käe-ø   all   *tegutse-nud,   suut-is   aprilli-s*
hand-GEN   under.LOC   operate-PST.PTCP   can-PST.3SG   April-INE
viimaks   ülejäänud   tööstuse-le   järele jõud-a …
finally   other   industry-ALL   catch up-INF
Lit. Chrysler, who has been working under the hand of the Italian car industry's
big name Fiat SpA for over a year now, finally managed to catch up with the rest
of the industry in April.
'Chrysler, who has been working under the Italian car industry´s big name Fiat
SpA for over a year now, finally managed to catch up with the rest of the
industry in April.' [lhv.e24.ee]

(102) *EBSi-ø   ja   ETI-ø   uurija-d   küsitle-si-d   kokku   25*
EBS-GEN   and   ETI-GEN   researcher-PL   inquire-PST-3PL   total   25
omanikku-ø   ja   tippjuhti-ø,   kelle-l   on   hea
owner-PRT   and   top manager-PRT   who-ADE   be.3SG   good
üldistusvõime   ja   ka   hulgaliselt   kogemus-i,   se-da   nii
generalisation   and   also   in quantities   experience-PL.PRT   this-PRT   so
*omaniku-ø   kui   tippjuhi-na,   nii   **kodu-ø   ja   välismaise-ø***
owner-GEN   as   top manager-ESS   as   home-GEN   and   foreign-GEN
kapitali-ø   käe-ø   all.
capital-GEN   hand-GEN   under.LOC
Lit. Researchers from EBS and ETI questioned altogether 25 owners and top
executives who are good at generalizing and who have a lot of experience both
as owners as well as top executives, under the hand of domestic as well as
foreign capital.
'Researchers from EBS and ETI questioned altogether 25 owners and top
executives who are good at generalizing and who have a lot of experience both
as owners as well as top executives, working with domestic as well as foreign
capital.' [www.director.ee]

The data show that the collective PNs only appear in examples where *käe all*
(hand+under) is a complex unit or a hybrid form. Table 4 shows that 58
instances out of 64 of the collective PNs occur with complex units. Whereas the
collective PNs are equally possible with both – examples that express

MENTAL CONTROL (as in (100)) as well as the PHYSICAL CONTROL (103). On six occasions, the collective PNs occur with hybrid forms (104).

(103) *Toitu   valmista-si-d   7   **meeskonda-ø, kelle-ø   käe-ø***
      food   make-PST-3PL   7   team-PRT   who-GEN   hand-GEN
      all   valmi-si-d   maitsva-d   boulanii-piruka-d   porrulaugu-ø,
      under.LOC   ripen-PST-3PL   tasty-PL   boulanii-pie-PL   leek-GEN
      kartuli-ø   ja   kanatäidise-ga,   pistaatsiabaklavaa,
      potatoe-GEN   and   chicken filling-COM   pistachio baklava
      meekana   ja   muu-d   hõrgutise-d.
      honey chicken   and   other-PL   dainty-PL
      Lit. Food was prepared by 7 teams under whose hand tasty boulan pies with leek, potato and chicken filling, pistachio baklava, honey chicken and other dainties became ready.
      'Tasty food like boulan pies with leek, potato and chicken filling, pistachio baklava, honey chicken and other dainties were prepared by 7 teams.' [maailmakool.ee]

(104) ***Kompartei-ø   juhtiva-ø   käe-ø   all***
      Communist party-GEN   controlling-GEN   hand-GEN   under.LOC
      ette võe-tud   üritus,   mis   lõppe-s   truualamliku-ø
      embark upon-PST.PTCP   event   what   end-PST.3SG   humble-GEN
      läkituse-ga   NLKP-ø   19.   parteikonverentsi-ø   delegaati-de-le,
      missive-COM   NLKP-GEN   19.   party conference-GEN   delegate-PL-ALL
      osutu-s   siiski   esimese-ks   vaba-ma-ks   foorumi-ks.
      turn out-PST.3SG   still   first-TRL   free-COMP-TRL   forum-TRL
      Lit. The event undertaken under the leading hand of the Communist party, which ended with a humble missive to the delegates of the 19th NLKP party conference, still turned out to be the first more liberal forum.
      'The event, which was controlled by the Communist party, ended with a humble missive to the delegates of the 19th NLKP party conference and still turned out to be the first more liberal forum.' [bhr.balanss.ee]

The PNs referring to humans and God may occur with all the structures. As the most common type of PN, the animate PNs make up most of the usages among the complex units (98%) and the freely combined phrases (99%).

The hybrid forms, while most commonly (68%) used with human PNs (see example (105)), are the most common structure among the PNs that refer to God (see example (106)) – 19 out of the total 25 examples of PNs that refer to God occur with this structure. Such usages pertain to religious contexts and have been around since at least the 17th century (see section 4.8.2.1).

(105) *Meie-ø   erakonna-d   on   arene-nud   peamiselt*
we-GEN   party-PL   be.3PL   develop-PST.PTCP   mainly
nendesama-de   juhti-de   kindla-ø   käe-ø   all,
these same-PL.GEN   leader-PL.GEN   firm-GEN   hand-GEN   under.LOC
kes   ol-i-d   nn   Asutava-te   Isakes-te
who   be-PST-3PL   so-called   Founding-PL.GEN   Father-PL.GEN
hulgas. [www.tlu.ee]
among
Lit. Our party has developed mainly under the the firm hand of those leaders
who were among the so called founding fathers.
'Our party has developed mainly led by those who were among the so called
founding fathers.' [www.tlu.ee]

(106) **_Jumala-ø   armastava-ø   käe-ø   all_**   *juhi-t-i*
God-GEN   loving-GEN   hand-GEN   under.LOC   conduce-IMPS-PST
minu-ø   pere   Pärnu-st   Tallinna-ø   ela-ma.
I-GEN   family   Pärnu-ELA   Tallinn-ILL   live-SUP
Lit. Under the loving hand of God, my family was led to live in Tallinn instead
of Pärnu.
'God's loving hand led my family to live in Tallinn instead of Pärnu.'
[www.advent.ee]

Thus, it was observed that *käe all* (hand+under) takes PNs that refer to three
semantic classes – human, God, and collectives. While the human PNs are
clearly the most common semantic class in all the structures, the collective PNs
only occur with complex units and hybrid forms. This is rather expected, as the
collective PNs are semantically incompatible with the simple structure. Thus,
their use with collective PNs is considered to indicate extension of the complex
postpositional *käe all* (hand+under).

4.5.1.2. *Külje all* (side+under)

As *külje all* (side+under) is also productively used to refer to object parts, the
semantic classes of PN are more diverse. The data in Table 5 shows that *külje
all* (side+under) co-occurs with PNs that belong to five semantic classes. The
PNs may refer to animate beings (humans and animals), collectives, (artificial)
objects, natural objects, or regions.

Table 5. The distribution of the semantic classes of the PN among the examples where *külje all* (side+under) occurs as a freely combined phrase, a complex unit, and a hybrid form

Semantic class of the PN	Total	Free	Unit	Hybrid
Animate	170 (7%)	55 (19%)	115 (5%)	0
Collective	52 (2%)	0	52 (2%)	0
Artificial object	176 (6%)	171 (59%)	5 (0%)	0
Natural object	82 (3%)	65 (22%)	16 (1%)	1 (13%)
Region	2083 (81%)	0	2076 (92%)	7 (88%)
Total	**2563 (100%)**	**291 (100%)**	**2264 (100%)**	**8 (100%)**

In this case, region is the most common PN class – 81% of the PNs of *külje all* (side+under) refer to a region (as in (107)). The PNs that refer to animate entities or collectives, as well as artificial and natural objects were used in 2–7% of the cases.

(107) *sihuke   küsimus   et   kus   **tallinna-ø**   **külje-ø**   **all***
      this kind   question   that   where   Tallinn-GEN   side-GEN   under.LOC
      see   filmi-tud   on?
      this   film-PST.IMPS.PTCP   be.3SG
      Lit. I was just wondering that where under the side of Tallinn has this been
      filmed.
      'I was just wondering where in the neighbourhood of Tallinn it was filmed.'
      [tv.motors24.ee]

The data suggests that PNs that refer to regions or collectives only occur with cases of *külje all* (side+under) that behave as a complex unit (as in (107)), whereas collective PNs occur in examples that express physical closeness (108) and mental closeness (109). The few abstract uses that were found in the data are coded as collective, because they are not frequent enough to form a distinct group. For instance in (110), where *külje all* (side+under) is preceded by the abstract notion *tähtaeg* 'deadline', it expresses the mental closeness of the coming deadline.

(108) *Transnistria   taha-b   Vene-ø   rakett-e   **NATO-ø***
      Transnistria   want-3SG   Russia-GEN   rocket-PL.PRT   NATO-GEN
      külje-ø   alla   *paiguta-da*
      side-GEN   under.LAT   deploy-INF
      Lit. Transnistria wants to deploy Russian missiles under the side of NATO.
      'Transnistria wants to deploy Russian missiles in the neighbourhood of NATO.'
      [bhr.balanss.ee]

(109) *Arvesta-des   filmitööstuse-ø   eripära-ø,   ole-ks   väikese-s*
consider-GER   film industry-GEN   peculiarity-PRT   be-COND   small-INE
Eesti-s   filmistuudio-l   majanduslikult   paslik   toimeta-da
Estonia-INE   film studio-ADE   economically   germane   work-INF
<u>telefirma</u>-ø   **külje-ø**   **all**.
television company-GEN   side-GEN   under.LOC
Lit. Taking into account the specificity of the film industry, in small Estonia it wold make financial sense for a film studio to operate under the side of a television company.
'Taking into account the specificity of the film industry, in small Estonia it would make financial sense for a film studio to operate together with a television company.' [www.ekspress.ee]

(110) *Õigupoolest   ol-i   juba   õite   ärevakstegevalt   **tähtaja-ø***
in fact   be-PST.3SG   already   quite   discouragingly   deadline-GEN
külje-ø   alla   *see   kuupäev   nihku-nud   kui*
side-GEN   under.LAT   this   date   shift-PST.PTCP   when
lõ-i   sisse   seesama   enesesäilitusinstinkt   mis   min-d
kick-PST.3SG   in   the same   self-preservation instinct   what   I-PRT
Luiga-ø   ja   Roberti-ga   Ruhrgebieti-lt   pääst-is
Luiga-GEN   and   Robert-COM   Ruhrgebiet-ABL   save-PST.3SG
paar   aasta-t   tagasi. [www.zaum.ee]
couple   year-PRT   ago
Lit. In fact the date had shifted quite discouragingly under the side of the deadline when the same self-preservation instinct that saved me with Luiga and Robert from Ruhrgebiet a couple of years ago kicked in.
'In fact the date had shifted quite discouragingly close to the deadline when the same self-preservation instinct that saved me with Luiga and Robert from Ruhrgebiet a couple of years ago kicked in.'

The rest of the semantic classes may occur with both – the freely combined and the complex units. However, each semantic class has its preferences. Artificial objects (as in (111)) and natural objects (as in (112)) are mostly used with freely combined phrases. In such cases, however, it is almost impossible to distinguish between the simple and the complex structure because mostly the combination of the freely combined phrase also implicitly includes the sense of proximity. Such examples are analyzed as the freely combined phrases, because in such cases the literal interpretation is still available and more prevalent. Cases where the object part meaning is less prevalent and the aspect of proximity more important are analyzed as complex units. In such cases, the literal interpretation is usually ruled out by the distance between the LM and the TR, plus the size of the TR relative to the LM. For instance, in (113) the actual distance of the KIASMA and its size relative to the equestrian statue of Mannerheim excludes a literal interpretation of *külje all* (side+under).

(111) *Mehe-d   ja   tema   tul-i-d   oma-ø   alumiiniumpanni-ga*
man-PL   and   s/he   go-PST-3PL   own-GEN   aluminium pan-COM
meie-ø   jahi-ø   külje-ø   alla   nii,   et   see
we-GEN   yacht-GEN   side-GEN   under.LAT   so   that   this
parras-t   kriipi-s.
board-PRT   rasp-PST.3SG
'The men and she with her aluminium pan came under the side of our yacht so that it rasped the broadside.' [www.ekspress.ee]

(112) *Veel   19.   sajandi-ø   alguse-s   ol-i   Simla*
yet   19th   century-GEN   beginning-INE   be-PST.3SG   Simla
*väheasustatud   koht   **Himaalaja-ø   külje-ø   all**,   kus*
underpopulated   place   Himaalaja-GEN   side-GEN   under.LOC   where
p-ol-nud   suur-t   rohkem   kui   mets,   looma-d
NEG-be-PST.PTCP   large-PRT   more   than   forest   animal-PL
ja   tempel.
and   temple
Lit. Still in the 19h century Simla was an underpopulated place under the side of the Himalayas where there was nothing much but forest, animals and a temple.
'Still in the 19h century Simla was an underpopulated place by the side of the Himalayas where there was nothing much but forest, animals and a temple.' [www.tlu.ee]

(113) *Kiasma-t   on   kõige   enam   kritiseeri-tud   liialt*
Kiasma-PRT   be.3SG   most   more   criticise-PST.PTCP   excessively
*ekspressiivse-ø   vormi-ø   ning   kindral   **Mannerheimi-ø***
expressive-GEN   form-GEN   and   general   Mannerheim-GEN
***ratsaskulptuuri-ø   külje-ø   alla   tikkumise-ø   pärast**.*
equestrian sculpture-GEN   side-GEN   under.LAT   intrude-GEN   because of
Lit. Kiasma has been mostly criticized for an excessively expressive form and for intruding under the side of the equestrian sculpture of General Mannerheim.
'Kiasma has been mostly criticized for an excessively expressive form and for trying to get in with the equestrian sculpture of General Mannerheim.' [arhliit.ee]

Animate PNs also occur with both – the simple and complex structure (114). However, Table 5 shows that animate PNs are more inclined to occur with freely combined phrases (as in (115)) in that they make up a larger proportion of usages (19%). However, in the case of *külje all* (side+under), the animate PN is not as frequent as in other phrases. However, the complex units are prevalently (92%) formed with PNs that refer to region (107).

(114) *Aga   no   see   olene-b   tõesti   ka   inimeses-t –   mõni*
but   well   this   depend-3SG   really   also   person-PRT   some
*koli-b   kohe   **teise-ø   tüdruku-ø   külje-ø   alla**.*
move-3SG   just   another-GEN   girl-GEN   side-GEN   under.LAT

Lit. But it also depends on the person – some move under the side of another girl straight away.
'But it also depends on the person – some move in with another girl straight away.' [buduaar.ee]

(115) *nimelt   sõit-si-n   sügise-l   porise-ø   ilma-ga*
      expressly drive-PST-1SG autumn-ADE muddy-GEN weather-COM
      kopli-s   galoppi-ø   ja   selle-le   hobuse-le   ikka   meeldi-b
      paddock-INE gallop-PRT and   this-ALL   horse-ALL   still   like-3SG
      vähe   kiiremini   kapa-ta   ja   kurvi-s   lenda-s   hobune
      little   faster   gallop-INF   and   curve-INE   fly-PST.3SG   horse
      *külili,   mu-l   jä-i   jalg   **hobuse-ø**   **külje-ø***
      sideways   I-ADE   get stuck-PST.3SG   leg   horse-GEN   side-GEN
      ***alla**.*
      under.LAT
      'In the autumn, I was galloping in a muddy paddock and this horse sure likes to
      gallop fast and in a curve the horse fell sideways, my leg got stuck under the side
      of the horse.' [www.lemmik.ee]

The hybrid forms that carry characteristics of both – the simple and the complex
structure – are very rare in the case of *külje all* (side+under). In the few such
examples, the PN refers almost without exception to region (116).

(116) *Mõlema-d   asu-vad   **Põlva-ø**   **võrupoolse-ø**   **külje-ø***
      both-PL   lie-3PL   Põlva-GEN   Võru sided-GEN   side-GEN
      ***all**   Rosma-ø   küla-s   Päkamäe-ø   kõrval.*
      under-LOC   Rosma-GEN   village-INE   Päkamäe-GEN   next to
      Lit. Both are situated under the Võru sided side of Põlva, in the village of
      Rosma, next to Päkamäe.
      'Both are situated near Põlva, towards Võru, in the village of Rosma, next to
      Päkamäe.' [rosma.edu.ee]

Thus, it was observed that the PNs of *külje all* (side+under) are rather diverse.
This is connected with the fact that the body part *külg* 'side' is also productively
used to refer to object parts. The data show that the PNs that refer to animate
beings, artificial objects, and natural objects, can be used with both structures,
but PNs that refer to collectives and regions are only used with the complex
units (and hybrid forms). Both of these PNs are also semantically incompatible
with the simple structure. Thus, in this case, both of these classes are considered
to indicate extension of *külje all* (side+under) as a complex unit.

4.5.1.3. *Selja taga* (back+behind)

Selja taga (back+behind) occurs with PNs that belong to three semantic classes –
animate being, object, and collective (Table 6). *Selja taga* (back+behind) pre-
fers the animate PNs, which in this case make up 90% of the data. The animate
PNs refer mostly to human beings. However, in some cases they also involve

PNs that refer to animals[50] or God. As the body part term *külg* 'side' can also be used to refer to object parts, PNs referring to objects form a distinct group of PNs. However, as such usages are uncommon, e.g. this group makes up less than 1% of the data. The collective PNs make up a modest 9% of the data. Thus, even though collective PNs do occur with *selja taga* (back+behind), they are quite rare.

Table 6. The distribution of the semantic classes of the PN among the examples where *selja taga* (back+behind) occurs as a freely combined phrase, a complex unit, and a hybrid form

Semantic class of the PN	Total	Free	Unit	Hybrid
Animate	3,312 (90%)	2,124 (98%)	1,173 (79%)	15 (79%)
Object	36 (1%)	34 (2%)	2 (<1%)	0
Collective	342 (9%)	20 (1%)	318 (21%)	4 (21%)
Total	**3,690 (100%)**	**2,178 (100%)**	**1,493 (100%)**	**19 (100%)**

The distribution of the PN classes among the freely combined phrases, the complex units and the hybrid forms shows that animate PNs occur with all these structures (as in (117)–(119)), but the animate PNs make up the largest group (98%) among the simple structure.

(117) *Ta   seis-is   **Aada-ø   selja-ø   taha**,*
      s/he   stand-PST.3SG   Aada-GEN   back-GEN   behind.LAT
      pan-i   käe-ø   Aada-ø   õla-le   ja
      put-PST.3SG   hand-GEN   Aada-GEN   shoulder-ALL   and
      vaata-s   laua-ø   poole.
      look-PST.3SG   table-GEN   to
      Lit. S/he stood behind the back of Aada, put his/her hand on Aada's shoulder and looked towards the table.
      'S/he stood behind Aada, put his/her hand on Aada's shoulder and looked towards the table.' [www.poogen.ee]

---

[50] *Selg* (back) as a body part of an animal is not necessarily associated with the posterior region but may generate function words referring to the superior region instead (Heine 1997, Svorou 1994). Therefore, the use of a PN referring to an animal could also be considered as an extension because the phrase *koera selja taga* (behind the dog's back) would not make sense given dogs are often in an upright position. However, the utterance *istuva koera selja taga* 'behind the back of the sitting dog' the can be interpreted as bearing a literal meaning, as there would be no mismatch between the body part term and the postposition. However, the corpus data, does not allow me to specify such details. Thus, I do not differentiate between PNs that refer to humans and animals.

(118) *Kui* **_oma-ø_** **_abikaasa-ø_** **_selja-ø_** **_taga_** *uus* *juba*
if own-GEN spouse-GEN back-GEN behind.LOC new already
salaja *välja* *vali-ta-kse,* *siis* *sellis-te-st* *inimes-te-st*
secretly out sort-IMPS-PR then this kind-PL-ELA people-PL-ELA
ma *suur-t* *ei* *pea-ø.*
I large-PRT NEG respect-CONNEG
Lit. If behind their spouse's back one is choosing a new one, I do not think much of people like this.
'I do not think much of people who choose a new partner behind the back of their spouse.' [www.maaleht.ee]

(119) *Iseäranis* *ajaloolis-te* *kultuuriväärtus-te-ga* *hoone-ø*
particularly historical-PL.GEN cultural value-PL-COM building-GEN
puhul *p-ole* *vastuvõetav,* *et* *ostja* *soovi-b*
in case of NEG-be acceptable that buyer wish-3SG
anonüümse-ks *jää-da* *puge-des* *advokaat* **_Viktor_** **_Kaasiku-ø_**
anonymous-TRL stay-INF creep-GER advocate Viktor Kaasik-GEN
laia-ø **_selja-ø_** **_taga_** *peitu.*
wide-GEN back-GEN behind.LAT into hiding
Lit. Especially in the case of historical buildings of cultural value, it is not acceptable that the buyer wishes to remain anynomous creeping to hide behind the wide back of the attorney Viktor Kaasik.
'Especially in the case of historical buildings of cultural value it is not acceptable for the buyer to wish to remain anynomous hiding behind the attorney Viktor Kaasik.' [larko.kolhoos.ee]

The PNs that refer to objects are mostly (34 instances out of 36) confined to the freely combined units (see example (120)), but in rare cases (less than 1%) they are used to complement the complex units. For instance, in (see example (121)) *selja taga* (back+behind) is used to express the ordinal relationship of camera K-5 II relative to other cameras.

(120) *Ning* *kes* *jää-b* *siis* *süüdi,* *kui* *sinine* *auto* *hakka-b*
and who stay-3SG then guilty if blue car start-3SG
sõit-ma **_rohelise-ø_** **_auto-ø_** **_selja-ø_** **_taha,_** *enda-l*
drive-SUP green-GEN car-GEN back-GEN behind.LAT own-ADE
sama-l *aja-l* *vasaku-le* *suund* *sees* *ja* *sama-l*
same-ADE time-ADE left-ALL blinker in and same-ADE
aja-l *roheline* *hakka-b* *pööra-ma?*
time-ADE green start-3SG turn-SUP

Lit. And who gets the blame if the blue car starts driving behind the back of the green car, at the same time indicating a turn to the left and at the same time the green one is starting to make a turn.
'And who gets the blame if the blue car starts driving behind the green car, at the same time indicating a turn to the left and at the same time the green one is starting to make a turn.' [www.auto24.ee]

(121) *Selle-ø   tulemuse-ga   jää-vad   **K-5 II-ø   selja-ø   taha***
this-GEN   result-COM   stay-3PL   K-5 II-GEN   back-GEN   behind.LAT
kõik   APS-C   ehk   poolkaadersensori-ga   peegelkaamera-d
all   APS-C   also known as   half frame sensor-COM   mirror camera-PL
ning   lisaks   ne-i-le   ka   veel   nii   mõne-d-ki
and   also   they-PL-ALL   also   more   so   some-PL-CL
täiskaadersensori-ga   profikaamera-d … .
full frame sensor-COM   professional camera-PL
Lit. With this result, all APS-Cs, also known as half frame sensor reflex cameras
as well as quite a few full frame sensor professional cameras remain behind the
back of K-5 II.
'This results leaves all APS-Cs, also known as half frame sensor reflex cameras,
as well as quite a few full frame sensor professional cameras behind the K-5 II.'
[blog.photopoint.ee]

In case of *selja taga* (back+behind), the collective PNs may occur with the
complex structure as well as the simple structure. However, the PNs referring to
collectives are more likely to occur with the complex structure (318 out of 342
instances). The collective PNs make up 21% of all the usages as complex post-
positions. This is considerably larger amount than the other phrases, where the
proportion of collective PNs remained around a few percents (except for *kaela
peal* (neck+on) see section 4.5.1.5). It is possible to use the collective modifier
with the simple structure because of the semantic and functional closeness
between the free combination of a body part term and a simple postposition and
the complex unit that expresses the locative function (see section 4.3.3). Thus,
such examples (as in (122)) may also be considered as in-between cases of the
simple and complex structure.

(122) *Raskekuulipilduja,   otse   **meie-ø   jao-ø   selja-ø***
heavy machine gun   straight   we-GEN   division-GEN   back-GEN
***taga,*   *vasta-s*   *pikka-de*   *valangu-te-ga*   *ja*
behind.LOC   answer-PST.3SG   long-PL.GEN   outpouring-PL-COM   and
paist-is,   et   punas-te   tuli   ol-i   selle-le
seem-PST.3SG   that   red-PL.GEN   fire   be-PST.3SG   this-ALL
suuna-tud.
direct-PST.PTCP
Lit. The heavy machine gun, directly behind the back of our our division,
replyed with long ourpurst and it seemed that the red's fire was directed at it.
'The heavy machine gun, directly behind our division, replyed with long
ourpurst and it seemed that the red´s fire was directed at it.' [www.virtsu.ee]

In a small amount of the examples (19) where *selja taga* occurs as a hybrid
form, the PN refers mostly to humans (see example (119)), but sometimes also
to collectives (see example (123)).

(123) *Just   seal,   suure-ø   **Matj-Rossiija-ø**   **kaitsva-ø**   **selja-ø***
right   there   big-GEN   Mother Russia-GEN   protective-GEN   back-GEN
***taga**,   väikene   Eesti   NSV   enda-le   turvalise-ø   koha-ø*
behind.LOC   small   Estonia   SSR   own-ALL   safe-GEN   place-GEN
leid-is-ki.
find-PST.3SG-CL
'Right there, behind the protective back of great Mother Russia is where the little
Estonian SSR found itself a safe place.' [www.epl.ee]

Thus, although the collective PNs occur with the freely combined phrases, the
complex units and the hybrid forms, it is prefers the complex unit. Thus, it is
also considered to indicate extension. However, as mentioned in section 4.3.3,
as a complex item *selja taga* (back+behind) occurs in four different functions –
SPACE-TIME, COVERTNESS, SUPPORT, and CONCEALMENT. In the
following, the distribution of PN lemmas will be observed among these func-
tions separately.

Table 7. The distribution of the semantic classes of the PN among the examples where
selja taga (back+behind) expresses SPACE-TIME, COVERTNESS, SUPPORT, and
CONCEALMENT

Semantic class of the PN	Total of Unit	SPACE/ TIME	COVERT- NESS	CONCEAL- MENT	SUPPORT
Animate	1,105 (74%)	224 (70%)	316 (79%)	160 (56%)	405 (83%)
God	68 (5%)	2 (1%)	0	65 (23%)	1 (<1%)
Object	2 (<1%)	1 (<1%)	0	0	1 (<1%)
Collective	318 (21%)	92 (29%)	84 (21%)	60 (21%)	82 (17%)
Total	**1,493 (100%)**	**319 (100%)**	**400 (100%)**	**285 (100%)**	**489 (100%)**

Table 7 shows that the PNs referring to God make up 23% of all the examples
of complex postpositional *selja taga* (back+behind) when it is used to express
CONCEALMENT (as in example (124)); in other functions, such PNs do not
exist at all or are extremely rare (1% of the cases).

(124) *Kuigi   Eesti   asu-b   seismiliselt   **vanajumala-ø**   **selja-ø***
although   Estonia   lie-3SG   seismologically   old god-GEN   back-GEN
***taga**,   ei   õnnestu-ø   ka   Eesti-s   pisara-i-d*
behind.LOC   NEG   success-CONNEG   also   Estonia-INE   tear-PL-PRT
ja   ver-d   pildi-s   välti-da.
and   blood-PRT   picture-INE   ignore-INF
Lit. Although Estonia is seismologically situated behind the back of God, it is
still not possible to avoid tears and blood in Estonia.
'Although Estonia is situated in a seismologically stable place, it is still not
possible to avoid tears and blood.' [www.ekspress.ee]

Due to the tendency for it to co-occur with PNs that refer to God, the amount of animate lemmas in the function CONCEALMENT is a little smaller compared to other functions. The amount of animate PNs reaches just above half of the usages (56%) of *selja taga* (back+behind) expressing CONCEALMENT, whereas in other functions it varies between 70–83%.

The collective PNs make up roughly $^1/_5$ of all the uses of *selja taga* (back+behind) as a complex postposition. Table 7 shows that the proportion of the collective PNs reaches almost 30% in case of SPACE/TIME. However, it must be noted that within this pattern, the collective PNs mostly occur when *selja taga* (back+behind) expresses the ordinal relationship, because in such examples *selja taga* (back+behind) is often used to describe team sports results (as in (126)) (see section 4.3.3). The amount of collective PNs is a little smaller in case of SUPPORT (17%), because in case of other functions the collective PNs make up 21% of the uses as a complex postposition (as in (125)).

(125) *Palju-d   riigi-d   on   en-d   **USA-ø**   **selja-ø***
       many-PL   country-PL   be.3PL   own-PRT   USA-GEN   back-GEN
       ***taha**   peitnud*. [www.vm.ee]
       behind.LAT   hide-PST.PTCP
       Lit. Many countries have hid themselves behind the back of the USA.
       'Many countries have hid behind the USA.' [www.vm.ee]

(126) *Suur-te-st   spordiriiki-de-st   jää-vad   **Eesti-ø***
       big-PL-ELA   sport country-PL-ELA   stay-3PL   Estonia-GEN
       ***seljataha**   näiteks   Soome,   Argentina   ja   Türgi*.
       back.behind.LAT   for example   Finland   Argentina   and   Turkey
       Lit. Of the great countries of sport, for example Finland, Argentina and Turkey remain behind the back of Estonia.
       'Of the great countries of sport, for example Finland, Argentina and Turkey are left behind by Estonia.' [london.postimees.ee]

Among the examples of *selja taga* (back+behind) there are instances where the PN refers to an abstract entity. Because of their overall infrequency, such usages have been coded as 'collective' as in the case of the other phrases analyzed above. Such usages are present in all of the functions of the complex unit *selja taga* (back+behind) (127)–(130).

(127) *Kui   inimese-d   saa-vad   rikka-ks,   haritu-ks   ja   seksuaalselt*
       if   people-PL   become-3PL   rich-TRL   educated-TRL   and   sexually
       vabastatu-ks,   siis   jäta-vad   na-d   irratsionaalse-d   usu-d
       liberated-TRL   then   leave-3PL   they-PL   irrational-PL   belief-PL
       *ja   **muu**   **jabura-ø**   seljataha*.
       and   other   harebrained-GEN   back.behind.LAT
       Lit. Once people become rich, educated and sexually liberated, they leave irrational beliefs and other lunacies behind their back.
       'Once people become rich, educated and sexually liberated, they leave behind irrational beliefs and other lunacies.' [syndikaat.ee]

(128) *Viies   küsimus: "Kuidas   Teie   keskkonnaministri-na   hinda-te*
fifth   question   How   You   environment minister-ESS   evaluate-2PL
*kavatsus-t   paiguta-da   tuumajaam   **avalikkuse-ø   selja-ø***
intent-PRT   deploy-INF   nuclear power station   public-GEN   back-GEN
taga   *tundliku-le   ja   väärtusliku-le   looduskeskkonna-ø*
behind.LOC   sensitive-ALL   and   valuable-ALL   nature environment-GEN
ala-le   ning   põlise-sse   inimasustuse-ø   piirkonda-ø?"
area-ALL   and   indigenous-ILL   human settlement-GEN   area-ILL
'The fifth question: "How do you as the Minister of Environment feel about the
intention to deploy a nuclear power station behind the back of the public onto a
sensitive and valuable area of natural environment and indigenous human
settlement?"' [www.riigikogu.ee]

(129) *... siis   ei   saa-ø   peitu-da   ka   **selle-ø**   **argumendi-ø***
then   NEG   can-CONNEG   hide-INF   also   this-GEN   argument-GEN
selja-ø   taha.
back-GEN   behind.LAT
Lit. Then you cannot hide behind the back of that argument either.
'Then you cannot hide behind that argument either.' [rahvahaal.delfi.ee]

(130) ***Üleilmse-ø   ettevõtmise-ø   "Teeme Ära Maailmakoristus 2012" selja-ø***
global-GEN   endeavor-GEN   "Teeme Ära Maailmakoristus 2012"   back-GEN
taga   *on   juba   nii   ÜRO   kui   ka   maailmanime-ga*
behind.LOC   be.3SG   already   so   UN   if   also   world famous-COM
ajakiri   Economist,   viimas-te-l   andme-te-l   ka   Maailmapank
magazine   Economist,   last-PL-ADE   data-PL-ADE   also   IBRD
ja   isegi   NATO.
and   even   NATO
Lit. Behind the back of the global endeavour 'Teeme Ära Maailmakoristus 2012'
there are already the UN as well as the world famous magazine the Economist;
according to the latest data even World Bank and NATO.
'The global endeavour "Teeme Ära Maailmakoristus 2012" is already backed by
UN as well as the world famous magazine the Economist; according to the latest
data even by World Bank and NATO.' [vikerraadio.err.ee]

As a complex unit *selja taga* (back+behind) was associated with collective PNs
more often than the other studied phrases (except for *kaela peal* (neck+on), see
section 4.5.1.5). Although the collective PNs may also occur to a certain extent
with the freely combined phrases, the collective PNs have a clear preference
towards the complex structure. Thus, the data suggests that *selja taga*
(back+behind) as a complex item has extended beyond human reference.

4.5.1.4. *Käe kõrval* (hand+beside)

Käe kõrval (hand+beside) takes three types of PNs – animate, collective, and
abstract; however, the latter two are extremely rare, occurring only once (see

Table 8). Thus, it seems that *käe kõrval* (hand+beside) is almost exclusively (98%) used with animate PNs, which in this case only refer to human beings.

Table 8. The distribution of the semantic classes of the PN among the examples where *käe kõrval* (hand+beside) occurs as a freely combined phrase, a complex unit, and a hybrid form

Semantic class of PN lemma	Total	Free	Unit	Hybrid
Animate	124 (98%)	5 (100%)	117 (98%)	2 (100%)
Collective	1 (1%)	0	1 (1%)	0
Abstract	1 (1%)	0	1 (1%)	0
Total	126 (100%)	5 (100%)	119 (100%)	2 (100%)

When we observe the distribution of PN lemmas among the freely combined phrases, complex units and hybrid forms, we see that both the collective PN (131) and the abstract PN (132) are used in examples where *käe kõrval* (hand+beside) behaves as a complex postposition.

(131) ***Renault-ø   käe-ø   kõrval***   *kõndiv   Rumeenia-ø   päritolu*
Renault-GEN   hand-GEN   beside.LOC   walking   Romania-GEN   origin
Dacia   on   õige   pea   oma-ø   ülisoodsa-t,
Dacia   be.3SG   pretty   soon   own-GEN   extremely favorable-PRT
lihtsa-t   ja   vastupidava-t   mudelivaliku-t
easy-PRT   and   durable-PRT   choice of model-PRT
laienda-ma-s. [www.delfi.ee]
broaden-SUP-INE
Lit. The Romanian originating Dacia walking beside the hand of Renault is about to broaden its extremely favourably priced, simple and durable selection of models very soon.
'The Romania-originating Dacia walking beside Renault is about to broaden its extremely favourably priced, simple and durable selection of models very soon.'

(132) *Kasvatus   aga   on   kõndi-nud   ühiskondlikult*
education   but   be.3SG   walk-PST.PTCP   socially
võimuka-te   väärtus-te   käekõrval.
powerful-PL.GEN   value-PL.GEN   hand.beside.LOC
Lit. Education, however, has been walking beside the hand of socially powerful values.
'Education, however, has been walking in hand with socially powerful values.' [opleht.ee]

The rest of the examples where *käe kõrval* (hand+beside) behaves as complex unit occur with animate (i.e. human) PNs (98%), which may occur in both the functions BESIDE (133) and ACCOMPANIMENT (134). The few examples that represent the simple structure and the hybrid forms only occur with human PNs.

(133) *Lisaks   veel   muidugi   see,   kui   tähista-t-i   võidupäeva-ø*
also   more   of course   this   when   celebrate-IMPS-PST   Victory Day-PRT
*ja   Helmut   **ema-ø**   **käe-ø**   **kõrval**   tõrviku-te*
and   Helmut   mother-GEN   hand-GEN   beside.LOC   torch-PL.GEN
valge-l   kalmistu-le   läks.
light-ADE   cemetery-ALL   go.PST.3SG
Lit. In addition of course when Victory Day was being celebrated and Helmut
went to the cemetery beside the hand of her mother in torchlight.
'In addition of course when Victory Day was being celebrated and Helmut went
to the cemetery beside her mother in torchlight.' [www.vorumaateataja.ee]

(134) *Ravimtaime-de-ga   sa-i   nüüdse-ks   juba   kuldse-sse*
herb-PL-COM   become-PST.3SG   current-TRL   already   golden-ILL
ikka-ø   jõud-nud   proua   sõbra-ks   oma-ø
age-ILL   reach-PST.PTCP   madam   friend-TRL   own-GEN
vanaema-ø   **käekõrval**.
grandmother-GEN   hand.beside.LOC
Lit. The madam, now in her golden years, became aquainted with medicinal
herbs beside the hand of her grandmother.
'The madam, now in her golden years, became aquainted with medicinal herbs
alongside her grandmother.' [arvamus.postimees.ee]

Thus, *käe kõrval* (hand+beside) is extremely rarely used with PNs that belong to
any other semantic class than human. However, two such cases – collective and
abstract – were found, both of which occurred with complex units. This result is
quite expected because these PNs are not semantically compatible with the
simple structure, which presumes a literal interpretation. Thus, these two cases
are considered to suggest extension.

4.5.1.5. *Kaela peal* (neck+on)

Kaela peal (neck+on) has three types of PNs – animate, object, and collective
PNs; in this case, the last is also quite rare – collective PNs make up only 10%
of the data. It can be observed in Table 9 that the body part related phrase is
most often (87%) modified by the animate PN. PNs that refer to objects are
extremely rare (3%).

Table 9. The distribution of the semantic classes of the PN among the examples where *kaela
peal* (neck+on) occurs as a freely combined phrase, a complex unit, and an hybrid forms

	Total	Free	Unit	Hybrid
Animate	63 (87%)	18 (90%)	44 (86%)	1 (100%)
Object	2 (3%)	2 (10%)	0	0
Collective	7 (10%)	0	7 (14%)	0
Total	72 (100%)	20 (100%)	51 (100%)	1 (100%)

Within the distribution of PN lemmas among the freely combined phrases, complex units and hybrid forms, collective PNs are confined to complex items (135), whereas there are also instances of abstract PNs which – due to extreme infrequency – are not presented as a separate class here (136). The PNs that refer to objects occur only with freely combined phrases (137). In this case too, the human PNs occur with the simple and the complex structure as well as the hybrid forms.

(135) *Ja   nii   jää-b-ki   arvutisõltlane   oma-ø   **vanema-te**
and   so   stay-3SG-CL   computer addict   own-GEN   parents-PL.GEN
***kaela-ø   peale   istu-ma*.**
neck-GEN   on.LAT   sit-SUP
Lit. And so it happens that a computer addict remains sitting on the neck of his/her parents.
'And so it happens that a computer addict remains dependent on his/her parents.'
[naistekas.delfi.ee]

(136) *50 000   maksumaksja-t   juures   ja   **sots**   **süsteemi-ø**   **kaela-ø***
50 000   taxpayer-PRT   more   and   social   system-GEN   neck-GEN
pealt   *ära   anna-b   juurde   lisaraha-ø,   mi-da   saa-b*
on.SEP   off   give-3SG   more   extra money-PRT   what-PRT   can-3SG
arsti-de   palgatõusu-ks   kasuta-da.
doctor-PL.GEN   advance in salary-TRL   use-INF
Lit. 50 000 more taxpayers and off the neck of the social system gives extra money that can be used for an advance in the salary for doctors.
'50 000 additional taxpayers who are now off the hands of the social system gives extra money that can be used for an advance in the salary for doctors.'
[www.delfi.ee]

(137) *... Enda-l   päris   pikalt   ol-nud   AYR   **mille-ø***
own-ADE   quite   long   be-PST.PTCP   AYR   what-GEN
kaela-ø   peal   *ol-i   Nr 003 ....*
neck-GEN   on.LOC   be-PST.3SG   Nr 003
' I myself have had for quite a long time an AYR which had no. 003 on its neck.'
[mootorratas.ee]

Thus, *kaela peal* (neck+on) is rarely used with any other PNs than animate, which occur with both – free and complex phrases. However, the rare cases of collective PNs are only used with complex units. In this case also, the collective PNs are semantically incompatible to modify the first component of the simple structure, and such usages are considered to indicate contextual expansion.

4.5.1.6. *Jalge all* (feet+under)

Table 10 shows that *jalge all* (feet+under) takes only two types of PNs – animate and collective PNs. In this case too, the collectives are quite rare as they make up just 3% of the examples where *jalge all* (feet+under) is preceded by a PN; examples with animate PNs dominate the data (97%). However, when the structures are observed individually, the collective PNs occur only with freely combined phrases.

Table 10. The distribution of the semantic classes of the PN among the examples where *jalge all* (feet+under) occurs as a freely combined phrase, a complex unit, and a hybrid form

Semantic class of the PN	Total	Free	Unit	Hybrid
Animate	366 (97%)	341 (97%)	24 (100%)	1 (100%)
Collective	9 (3%)	9 (3%)	0	0
Total	**375 (100%)**	**350 (100%)**	**24 (100%)**	**1 (100%)**

A closer look at such examples shows that these examples represent idiomatic expressions. For instance, in (138) *jalge all* (feet+under) is part of the expression *maa põleb jalge all* (lit. ground beneath one's feet is on fire 'the situation is (getting) dangerous')[51] and in (139), *jalge all* (feet+under) is part of the expression *kindel pind* (*on*) *jalge all* (lit. secure surface (is) beneath one's feet 'one has achieved security '[in some aspect])[52]. Although these examples are instance of non-literal use, the phrase *jalge all* (feet+under) does not behave as a holistic unit that carries a distinct meaning. Thus, such cases are considered to represent freely combined phrases that are part of larger idiomatic expressions that are also responsible for the extension of *jalge all* (feet+under) to collective PNs. Thus, the data suggests that as a complex unit, *jalge all* (feet+under) only takes animate PNs (as in (140)). This suggests that *jalge all* (feet+under) as a complex postposition has not extended beyond human reference.

(138) *Maa* **_Reformierakonna-ø_** **_jalg-e_** **_all_** *põle-b*.
       earth Reform Party-GEN foot-PL.GEN under.LOC burn-3SG
       Lit. The ground under the feet of the Reform Party is on fire.
       'The situation of the Reform Party is unstable.' [www.epl.ee]

---

[51] http://www.eki.ee/dict/ekss/index.cgi?Q=p%C3%B5leb+jalge+all&F=M (Accessed 11.01.2016)
[52] http://www.eki.ee/dict/ekss/index.cgi?Q=kindel+pind+jalge+all&F=M (Accessed 11.01.2016)

(139) *"Olukord   on   küll   endiselt   raske,   kuid pind*
       situation   be.3SG   although   still   complicated   but   surface
       Valio   Eesti   jalg-e   all   on   kindel   ning
       Valio   Eesti   foot-PL.GEN   under.LOC   be.3SG   stable   and
       põhjus-t   liigse-ks   paanika-ks   me   ei   näe-ø,"
       reason-PRT   excessive-TRL   panic-TRL   we   NEG   see-CONNEG
       lisa-s   ta.
       add-PST.3SG   s/he
       Lit. ″Although the situation is still complicated, the surface under the feet of
       Valio Eesti is stable and we see no reason for excessive panic, ″ s/he added.
       'Although the situation is still complicated, Valio Eesti is standing on solid
       ground and we see no reason for excessive panic, ″ s/he added.' [www.epl.ee]

(140) *Me   püüd-si-me   tee-de-lt   kõrvale,   põldu-de-le   põiga-ta,*
       we   try-PST-1PL   road-PL-ABL   beside   field-PL-ALL   dodge-INF
       *et   mitte   **inimes-te**   **jalg-e**   **alla***
       that   not   human-PL.GEN   foot-PL.GEN   under.LAT
       talla-tud   saa-da.
       trample-PST.PTCP   become-INF
       Lit. We tried to dodge off the roads, onto the fields, in order not to get trampled
       under the feet of people.
       'We tried to dodge off the roads, onto the fields, in order not to get trampled by
       people.' [www.advent.ee]

4.5.1.7. Summary and conclusions based on the distribution of the semantic classes of the PN lemma

Most of the studied phrases occur with PN lemmas referring to human beings[53], objects, collectives, and, to a lesser extent, abstract notions. In the case of all the phrases, the animate (mostly human) PNs make up the largest proportion and the collective PNs are rather modestly represented (1–9% of the examples). This suggests that based on the semantic class of the PN lemma, the expansion of the studied phrases is not particularly substantial.

However, in most cases the semantic classes have a preference towards the structures with which they occur. For instance, collective PNs are in most cases (*käe all* (hand+under), *külje all* (side+under), *käe kõrval* (hand+beside), *kaela peal* (neck+on)) used only with complex units or with in-between cases, and the PNs that refer to objects mostly (*külje all* (side+under), *selja taga* (back+ behind), *kaela peal* (neck+on)) occur with freely combined units. Both of these results are rather expected. Complex units tend to co-occur with collective PNs because they are lexicalized, the association with the body part meaning is fading, and, therefore, they can co-occur with such PNs that are semantically incompatible with the literal meaning. The body part term in most of the studied

---

[53] Or other animate beings that have body parts (animals) or beings that can be projected in the form of a human being (God).

phrases (except for *külje all* (side+under) and *kaela peal (*neck+on)) is not productively used as an object parts. Thus, it is not likely that such uses would be frequent among the complex items either.

The collective PNs are most common among the complex units of *selja taga* (back+behind) and *kaela peal* (neck+on), where such usages made up 21% and 14% of all the uses as a complex postposition respectively. In the case of other phrases, the collective PNs made up just 1–2% of the data. The data suggests that *jalge all* (feet+under) has not extended to be used with PNs that refer to collectives.

Human PNs, however, are extensively used with both the simple and the complex structure. This suggests that the complex postpositions have developed in the context of animate (human) PNs. Thus, the (human) collective PNs are the logical following step towards the more abstract uses, which have been made possible by a semantic change that must have taken place in the context of animate PNs. It was mentioned that the label 'Collective' also includes instances of PNs that refer to institutions or abstract notions, which can be considered as the next steps in the extension of complex postpositions. The further extension to other semantic classes is rather expected, as entering even more abstract contexts would suggest dissociating from the source form and its meaning, which is in line with the general principles of grammaticalization. Thus, the data suggests that the expansion of the PN lemmas follows this path:

ANIMATE (HUMAN) > (HUMAN) COLLECTIVE > INSTITUTION (> ABSTRACT)

The class 'abstract' is placed in brackets because there is no direct link between the classes INSTITUTION and ABSTRACT. The label ABSTRACT is also very general, which is connected with the fact that such examples are still quite rare and it is difficult to pinpoint the possible subtypes of this class. On the other hand, the presence of abstract PNs may also be an independent development, in the sense that the PN classes have not necessarily expanded via one path, but the already existing complex unit may also have expanded via supplementary paths.

The above path suggests that in cases where the first component of the body part related phrase is not productively used to refer to object parts, the object part stage is not vital for the development of the complex function words. Thus, such cases do not follow the model of grammaticalization of body part terms (Svorou 1994: 90; Heine 1997: 44), which suggests that the development of function words that originate from body part terms also pass through the stage of the object part. However, in this case the role of object part is marginal and the shift from the simple structure to the complex structure seems to have taken place in the context of human PNs. The reason being, that in case of the development of complex postpositions in Estonian, it is not considered to be the grammaticalization of body part terms in the usual sense, in which a single body part term develops into a spatial function word, but a more complex process that involves larger and more complex elements that have to develop a holistic

meaning, i.e. lexicalize, and begin to express abstract meanings. Thus, it seems that because the phrases are first lexicalized, they get to skip certain stages in the usual grammaticalization path, such as the object part and (in most cases) the spatial function word stage.

It is assumed that in most cases, the human based developmental path is connected with the semantics of the source form. For instance, according to van Pareren (2013: 100–101), in Mordvin the body part *hand* has not grammaticalized into an adposition that refers to relative spatial areas, but has developed into an adposition that refers to humans' living quarters. The adpositional use needs a reference to human beings and is not linked to inanimate objects (ibid.). This also applies for the Estonian *käe all* (hand+under), because *käsi* 'hand' is not used to refer to object parts at all. However, the data suggests that inanimate objects are not part of this grammaticalization path even when the body parts are used to refer to objects to some extent (as in case of *selg* 'back'). For example, there was only one example where an object part served as a complement to the spatio-temporal complex postposition. This suggests that it is unlikely that the spatio-temporal usages of *selja taga* (back+behind) have passed by the stage of object part but rather that the complex items have evolved in the context of animate beings.

Nevertheless, in the case of *külje all* (side+under) the usage of PN semantic classes was more diverse. Because *külg* (side) is also productively used to refer to object parts, it exhibits a more diverse use of PNs and less clear distribution of the PNs between the simple and complex structures. Namely, the data show that both – the human PNs and object (including the natural objects) PNs are used with the simple structure and the complex structure. However, PNs that refer to collectives and regions only co-occur with complex units. Similar to the analysis above, I suggest that the use of the complex items with collective PNs suggest extension based on the usages with human PNs. Likewise, the use of PNs that refer to regions seems to be an extension from uses in the context of object parts with an intermediate stage of natural objects. Thus, the data suggests that the extension of *külje all* (side+under) follows two paths:

A) ANIMATE (HUMAN) > COLLECTIVE > INSTITUTIONAL (> ABSTARCT)

B) OBJECT > NATURAL OBJECT > REGION

Based on the analysis of the contemporary data, it remains unclear which path, if either, has occurred earlier, and is, therefore, responsible for the development of the sense of proximity. Most likely the complex postpositional *külje all* (side+under) has developed in both of these contexts. However, based on the data, it seems that path B is the dominant path because the examples where *külje all* (side+under) takes PNs that are artificial objects, natural objects, or regions are considerably more frequent than PNs that refer to human or (human) collectives. This question is further discussed in the diachronic analysis (see section 4.8.3).

Thus, the data suggest that while the use of the studied phrases is not extensively diverse, there is evidence of extension of the phrases that is observable in the semantic classes of the PNs. Semantic classes and their distribution among the simple and the complex structure help to shed some light on the developmental paths of the complex postpositions. The evidence suggests that the phrases that are not productively used to refer to object parts have developed into complex postpositions in the context of human PNs, whereas the phrases that are productively used to refer to object parts likely have two parallel paths starting with uses with humans or (natural) objects.

4.5.2. Decategorialization of the studied phrases

In this section, I present an analysis of the non-agreement between the PN and the body part term. The studied phrases are mostly (except for *jalge all* (feet+under)[54]) lexicalised in the singular form, i.e. *kaela-Ø peal* (neck-SG.GEN + on), not *kaela-de peal* (neck-PL.GEN + on) (see section 4.1). However, the PN that precedes the phrase may be either singular, i.e. *ema-Ø kaela-Ø peal* (mother-SG.GEN neck-SG.GEN on), or plural, i.e. *vanema-te kaela-Ø peal* (parent-PL.GEN neck-SG.GEN on). The non-agreement in number suggests semantic incompatibility with the simple structure, where the phrase is interpreted literally. Such usages are here considered to be suitable to occur with the complex units where the phrases are lexicalised and the agreement with the body part term is not required any more in order to make sense semantically. This phenomenon indicates contextual expansion, which is considered to be the actualization of the reanalysis process, whereby the structural relations between the constituents of the utterance have been re-interpreted (see section 2.5.3.2).

In the following section it will be observed:

i   to what extent are plural PNs used with the studied phrases?
ii   what is the distribution of plural and singular forms among the PNs that occur with the freely combined phrases and complex units?

The distribution of the singular and plural forms among the simple and complex structures are not as clear-cut as for the semantic classes. Thus, where appropriate, I use the chi-square test to determine whether the distribution of the singular and plural forms is significantly different.

---

[54] The phrase *jalge all* (feet+under) is not discussed in this section because in this case, the possible categorization is not observable in terms of non-agreement.

4.5.2.1. *Käe all* (hand+under)

In this section, I analyze the non-agreement in number between the body part term *käsi* (hand) and the PN. The non-agreement is observed where *käe all* (hand+under) is used as a freely combined phrase, a complex unit, or a hybrid form. Table 11 shows the distribution of the singular and plural forms in the two structures and hybrid forms as absolute numbers and percentages.

Table 11. The distribution of singular and plural PNs among the instances of *käe all* (hand+under) as a freely combined phrase, a complex unit, and a hybrid form

Grammatical number of the PN	Total	Free	Unit	Hybrid
SG	3307 (79%)	85 (97%)	3152 (78%)	70 (89%)
PL	886 (21%)	2 (3%)	875 (22%)	9 (11%)
Total	**4193 (100%)**	**87 (100%)**	**4027 (100%)**	**79 (100%)**

Käe all (hand+under) mostly occurs with singular PNs (as in example (141)). Singular PNs make up 79% of the examples. However, the examples with plural PNs (as in example (142)) still represent a considerable amount of the examples at 21% (887 examples out of 4192).

(141) *Seal   jä-i   **Veiko   Hormi-ø   käe-ø   alla***
      there   leave-PST.3SG   Veiko   Horm-GEN   hand-GEN   under.LAT
      mingi   suure-m   ese,   mille-ø   väljatõstmise-ks   võe-t-i
      some kind   big-COMP   thing   what-GEN   lifting out-TRL   take-IMPS-PST
      taas   appi-ø   suure-m   tõstekott. [www.lounaleht.ee]
      again   help-ILL   big-COMP   lifting bag
      'There, some bigger object appeared under Veiko Horm's hand and to lift it out a bigger lift bag was used again.'

(142) *Se-da,   et   film   on   **amatööri-de   käe-ø   all***
      this-PRT   that   film   be.3SG   amateur-PL.GEN   hand-GEN   under.LOC
      valmi-nud,   või-b   näh-a   ka   inimes-te
      mature-PST.PTCP   might-3SG   see-INF   also   people-PL.GEN
      näitlemisoskuses-t.
      acting skill-PRT
      Lit. That the film has become ready under the hand of amateurs can also be seen from the people's acting skill.
      'That the film has been done by amateurs can also be seen from the people's acting skill.' [filmitalgud.ee]

The distribution of the singular and plural forms among the simple and the complex structure show that the plural forms are much more prominent among the complex units than among the freely combined forms. In the case of the former, the plural PNs (see example (142)) occur on 875 occasions (out of 4027

examples), i.e. the plural forms make up 22% of the uses with PNs. In the case of the freely combined phrases, only two examples with plural forms could be found (3% of the usages). Both of these cases (as in example (143)) represent a usage where the body part is used in a more general sense to refer to a human hand (c.f. example (44) in 2.5.3.2). The hybrid forms also occur with the plural PNs (as exemplified in (144)) as well as the singular PNs, but are mostly (89%) used with the latter.

(143) *Kui   kuvar   kõrvale   jät-ta,   on   usin   hiireke   **meie-ø***
       if   monitor   beside   leave-INF   be.3SG   diligent   mouse   our-GEN
       käe-ø   all   just   see   seade,   mi-da   arvuti-ga
       hand-GEN   under.LOC   right   this   device   what-PRT   computer-COM
       tööta-des   kõige   rohkem   kasuta-me. [arhiiv.koolielu.ee]
       work-GER   most   more   use-1PL
       'If we leave aside the monitor, it is the diligent mouse under our hand that is used most when working with a computer.' [arhiiv.koolielu.ee]

(144) *Palju-d   Eesti-ø   kunagise-d   vennasvabariigi-d   ning*
       many-PL   Estonia-GEN   has-been-PL   soviet country-PL   and
       lugematu   hulk   riik-e   üle   maailma-ø   ela-vad
       countless   amount   country-PL.PRT   over   world-GEN   live-3PL
       *endiselt   autoritaarse-te   **isakes-te   karmi-ø   käe-ø***
       still   authoritarian-PL.GEN   father-PL.GEN   tough-GEN   hand-GEN
       all   *ning   või-vad   demokraatia-t   vaid   une-s   näh-a.*
       under.LOC   and   can-3PL   democracy-PRT   only   sleep-INE   see-INF
       Lit. Many countries similar to Estonia in their background as well as a countless number of other countries all over the world still live under the strict hand of authoritarian fathers and can only dream of democracy.
       'Many countries similar to Estonia in their background as well as a countless number of other countries all over the world still live under authoritarian leaders and can only dream of democracy.' [www.ngo.ee]

Thus, the data shows that the plural PN is more prominent among the complex usages. This result is expected because in the case of the simple structure, the plural PN modifying the singular body part is semantically incompatible in most cases. The results are also statistically significant (*X*-squared = 23.8, *df* = 2, *p*-value = 0.01[55]). However, the effect size – which shows the strength of the correlation (Gries 2014: 5) – is quite weak (*V* = 0.07).

4.5.2.2. *Külje all* (side+under)

In the case of *külje all* (side+under), plural PNs are much less common than in the case of *käe all* (hand+under) analyzed above. The examples with plural PNs make up just 3% (84 out of 2563) of the data (see Table 12).

---

[55]   If the signicicance level is considered to be 0.05 for the humanities.

Table 12. The distribution of singular and plural PNs among the instances of *külje all* (side+under) as a freely combined phrase, a complex units and a hybrid form

Grammatical number of the PN	Total	Free	Unit	Hybrid
SG	2479 (97%)	270 (93%)	2201 (97%)	8 (100%)
PL	84 (3%)	21 (7%)	63 (3%)	0
Total	**2563 (100%)**	**291 (100%)**	**2264 (100%)**	**8 (100%)**

The distribution of the singular and plural forms among the simple and the complex structure shows that even though the plural forms are equally poorly represented among the simple and complex structures, the former is a little more inclined to occur with plural PNs. The examples with plural PNs (as exemplified in (145) and (146)) make up 3% (63 examples out of 2264) of the complex forms and 7% (21 examples out of 291) of the freely combined phrases. The freely combined phrases occur with plural PNs in cases where the body part is used to refer to the body part *side* in general (as in example (145) and in cases where *külje all* (side+under) is used to describe the relationship of a tall/high object relative to smaller object(s) (as in (147)). In latter cases, alternatively, the analysis as the complex structure is also possible, which is elicited by the plural PNs. The data suggests that the hybrid forms only took singular PNs.

(145) *"Jeesus,   na-d   on   otse   **meie-ø   külje-ø   all!"***
      Jesus   they-PL   be.3PL   straight   we-GEN   side-GEN   under.LOC
      hüüata-s   Ilona.
      exclaim-PST.3SG   Ilona
      Lit. "Jesus, they are straight under our side!" exclaimed Ilona.
      '"Jesus, they are right beside us!" exclaimed Ilona.' [ulmeajakiri.ee]

(146) *Sageli   ol-i-d   **nen-de   külje-ø   all**   ainult*
      frequently   be-PST-3PL   they-PL.GEN   side-GEN   under.LOC   only
      hein   ja   kuivanud   puulehe-d.
      hay   and   dried   leaf-PL
      Lit. Often, there were only hay and dried leaves under their sides.
      'Often, they only had hay and dried leaves under them.' [www.hambaarst.ee]

(147) *Nee-d   väikse-d   viiekorruselise-d   siin   suur-te   **üheksakordse-te***
      this-PL small-PL   five-storey-PL   here   big-PL.GEN   nine-storey-PL.GEN
      külje-ø   all   ei   kao-ø   kuhugi.
      side-GEN   under.LOC   NEG   disappear-CONNEG   nowhere
      Lit. These little five-storey [buildings] here under the side of the big nine-storey ones are going nowhere.
      'These little five-storey [buildings] here beside the big nine-storey ones are going nowhere.' [www.minaolinsiin.ee]

Thus, the data suggests that plural PNs are not particularly commonly used with *külje all* (side+under), and that the plural forms are not common in complex units either. In fact, the data suggests that the simple structure is more inclined to be used with the plural PNs. This result is statistically significant (X-squared $= 14.9$, $df = 1$, p-value $= 0.0001$) but the effect size suggests weak correlation ($V = 0.07$).[56]

4.5.2.3. *Selja taga* (back+behind)

The data show that plural PNs make up 18% (679 examples out of 3,690) of the data of *selja taga* (back+behind).

Table 13. The distribution of singular and plural PNs among the instances of *selja taga* (back+behind) as a freely combined phrase, a complex unit, and a hybrid form

Grammatical number of the PN lemma	Total	Free	Unit	Hybrid
SG	3,011 (82%)	1,871 (86%)	1,124 (75%)	16 (84%)
PL	679 (18%)	307 (14%)	369 (25%)	3 (16%)
Total	3,690 (100%)	2,178 (100%)	1,493 (100%)	19 (100%)

Although plural forms are present in simple and complex structures and hybrid forms, the plural forms are most common among the complex units (as in example (148)). The examples with plural PNs make up 25% (369 examples out of 1,124) of all the uses of *selja taga* (back+behind) as a complex unit. Among the freely combined phrases and hybrid forms, such cases make up 14% (307 examples out of 2,178) and 16% (3 examples out of 19) respectively. The freely combined phrases occur with plural PNs mostly in cases where the PN refers to LMs that are located closely together and facing the same direction so that the TR is in the same location relative to the both of them (149). In such cases, *selja taga* (back+behind) behaves similarly to a complex spatial function word, and the plural form contributes to the latter analysis.[57]

---

[56] This result may be a little magnified. For the statistical analysis, the freely combined phrases and the hybrid forms were merged into one category (non-complex structure) because the contingency table did not meet the requirements for the chi-squared test, which does not allow empty cells.

[57] In order to not over-interpret the data, such usages have been coded as instances of freely combined phrases.

(148) *Varem   või   hiljem   avasta-vad   tavalise-d   inimese-d,   et*
sooner   or   later   find-3PL   regular-PL   people-PL   that
Euroopa-ø   Liit   ja   Euroopa-ø   riiki-de   juhi-d
Europe-GEN   Union   and   Europe-GEN   country-PL.GEN   leader-PL
*on   **kohalik-e   elanik-e   selja-ø   taga***
be.3PL   local-PL.GEN   dweller-PL.GEN   back-GEN   behind.LOC
otsusta-nud,   et   moslemi-d   saa-vad   vabalt   koloniseeri-da
decide-PST.PTCP   that   muslim-PL   can-3PL   easily   colonise-INF
meie-ø   kontinendi-ø.
we-GEN   continent-GEN
'Sooner or later ordinary people will discover that the European Union and the leaders of the European countries have decided behind the backs of the local people that muslims can freely colonise our continent.' [arvamus.postimees.ee]

(149) ***Aheldatu-te   selja-ø   taga***   *asu-b   lõke,   mis*
chained-PL.GEN   back.GEN   behind.LOC   lie-3SG   log fire   what
valgusta-b   koobas-t.
light-3SG   cave-PRT
Lit. Behind the backs of the chained [people] there is fire that lightens the cave.
'Behind the chained [people] there is fire that lightens the cave.' [kalah.zzz.ee]

Thus, the data suggests that plural PNs are most common among the complex units. This result is statistically significant (*X*-squared = 66.62, *df* = 2, *p*-value < 0.001), however in this case the effect size also suggests a rather weak correlation (V = 0.13).

As the complex postpositional *selja taga* (back+behind) carries four functions – SPACE-TIME, COVERTNESS, SUPPORT, and CONCEALMENT, it makes sense to observe the distributions of the singular and plural PNs among these functions separately (Table 14).

Table 14. The distribution of singular and plural PNs among the functions of *selja taga* (back+behind)

Grammatical number of the PN	Total	Free	SPACE-TIME	COVERT-NESS	CONCEAL-MENT	SUP-PORT	Hybrid
SG	3,011 (82%)	1,871 (86%)	236 (74%)	292 (73%)	229 (80%)	367 (75%)	16 (84%)
PL	679 (18%)	307 (14%)	83 (26%)	108 (27%)	56 (20%)	122 (25%)	3 (16%)
Total	**3,690 (100%)**	**2,178 (100%)**	**319 (100%)**	**400 (100%)**	**285 (100%)**	**489 (100%)**	**19 (100%)**

Table 14 shows that the amount of plural PNs is largest in the case of COVERT-NESS (see example (148)), SPACE-TIME (150), and SUPPORT (151), where the plural forms make up 27% (108 examples out of 400), 26% (83 examples out of 26%), and 25% (122% out of 489) of the data. The amount of plural PNs is smaller when *selja taga* (back+behind) used to express CONCEALMENT (see example 152)), where the plural PNs make up 20% of the complex postpositions. This result is statistically significant (*X*-squared = 73.261, *df* = 5, *p*-value < 0.001) but the effect size suggest that the correlation is rather weak (V= 0.14).

(150) ***Eestlas-te***   ***selja-ø***   ***taha***   *jä-i*   *üks*
Estonian-PL.GEN   back-GEN   behind.LAT   leave-PST.3SG   one
konkurent.
competitor
Lit. One competitor was left behind the back of the Estonians.
'There was one competitor who came in behind the Estonians.' [www.vsport.ee]

(151) *Miks   eestimaalase-d   ei   tule-ø*   ***arsti-de***   ***selja-ø***
why   Estonian-PL   NEG   go-CONNEG   doctor-PL.GEN   back-GEN
taha   *üldstreigu-ø*   *näol,*   *kus   sa   Taliga*
behind.LAT   general strike-GEN   in form of   where   you   Taliga
ole-d   oma-ø   aü-ga
be-2SG   own-GEN   trade union-COM
Lit. Why don't Estonians come behind the backs of the doctors in the form of a general strike; where are you Taliga with your trade union?
'Why don´t Estonians support the doctors with a general strike; where are you Taliga with your trade union?' [www.maaleht.ee]

(152) *Poliitiku-d*   *poe-vad*   ***eksperti-de***   ***selja-ø***
Politician-PL   curry favor-3PL   expert-PL.GEN   back-GEN
taga. [www.vorumaateataja.ee]
behind.LOC
Lit. Politicians hide behind the backs of experts.
'Politicians hide behind expert opinions.' [www.vorumaateataja.ee]

4.5.2.4. *Käe kõrval* (hand+beside)

It can be observed in Table 15 that in the case of *käe kõrval* (hand+beside), plural PNs make up 12% (15 examples out of 126) of all usages with PNs. Thus, the use of plural PNs is not particularly common with *käe kõrval* (hand+beside). However, it should be noted the data sample is quite small compared to the phrases discussed above (see section 4.1).

Table 15. The distribution of singular and plural PNs among the instances of *käe kõrval* (hand+beside) as a freely combined phrase, a complex unit, and a hybrid form

Grammatical number of the PN	Total	Free	Unit	Hybrid
SG	111 (88%)	5 (100%)	106 (89%)	0
PL	15 (12%)	0	13 (11%)	2 (100%)
Total	**126 (100%)**	**5 (100%)**	**119 (100%)**	**2 (100%)**

When singular and plural forms are analyzed separately among the simple and complex structure, it can be observed that plural PNs do not co-occur with the free forms at all. Plural PNs occur with complex units on only 13 occasions (for example in (153)), which make up 11% of all uses of *käe kõrval* (hand+beside); the rest of the plural PNs (2 instances) occur with hybrid forms (as for example in (154)).

(153) *Järelikult   pea-ks   osa   töökasvatuse-st   ole-ma   üle*
        consequently   should-COND   part   work education-ELA   be-SUP   over
        kandu-nud   kooli-de-sse,   nii,   et   laps   õpi-ks
        spread-PST.PTCP   school-PL-ILL   so   that   child   learn-COND
        lasteaia-s   või   kooli-s   se-da,   mi-da   varem
        kindergarten-INE   or   school-INE   this-PRT   what-PRT   sooner
        loomulikul   teel   õpi-t-i   pere-s
        natural   way   learn-IMPS-PST   family-INE
        ema-de-isa-de   **_käe_**   **_kõrval_**.
        mother-PL.GEN-father-PL.GEN   hand-GEN   beside.LOC
        Lit. Therefore, part of the work education should be transferred to schools, so that a child would learn in the kindergarden or schools the things that used to be learnt more naturally beside the hand of mothers and fathers.
        'Therefore, part of the work education should be transferred to schools, so that a child would learn in the kindergarden or schools the things that used to be learnt more naturally alongside mothers and fathers.' [vikerraadio.err.ee]

(154) *Illuka-ø   Kooli-ø   van-i-ma   klassi-ø   õpilas-te-l*
        Illuka-GEN   School-GEN   old-SPL-GEN   grade-GEN   student-PL-ADE
        ol-i   vahva   võimalus   30. mai-l   toimu-nud
        be-PST.3SG   nice   opportunity   30. May-ADE   take place-PST.PTCP
        tutipäeva-ø   aktuse-le   sammu-da   esimese-ø
        dress-up day-GEN   public ceremony-ALL   step-INF   first-GEN
        *klassi-ø   **_õpilas-te_**   **_toetava-ø_**   **_käe-ø_**   **_kõrval_**.
        grade-GEN   student-PL.GEN   supporting-GEN   hand-GEN   beside.LOC
        Lit. On May 30, their last day of school, seniors of the Illuka school had a great opportunity to walk to the final ceremony besides the supporting hand of the firstgraders.
        'On May 30, their last day of school, seniors of the Illuka school had a great opportunity to walk to the final ceremony by the supporting firstgraders.' [www.delfi.ee]

In conclusion, it was observed that the plural PNs are not particularly frequently used with *käe kõrval* (hand+beside). The data samples did not include any instances of plural PNs being used with freely combined phrases. Thus, in case of *käe kõrval* (hand+beside), the plural PNs only occur with complex units and hybrid forms.[58]

4.5.2.5. *Kaela peal* (neck+on)

Despite the fact that *kaela peal* (neck+on) is one of the least frequent among the studied phrases and that there are only 72 examples where a PN precedes the phrase, plural PNs are rather frequent. It can be observed in Table 16 that *kaela peal* (neck+on) is preceded by a plural PN in 31% of the cases (22 examples out of 72).

Table 16. The distribution of singular and plural PNs among the instances of *kaela peal* (neck+on) as a freely combined phrase, a complex unit, and a hybrid form

Grammatical number of the PN	Total	Free	Unit	Hybrid
SG	50 (69%)	19 (95%)	30 (59%)	1 (100%)
PL	22 (31%)	1 (5%)	21 (41%)	0
Total	**72 (100%)**	**20 (100%)**	**51 (100%)**	**1 (100%)**

When the simple and the complex structure are observed separately, it is found that the plural PNs are even more frequent among complex units where such usages make up 41% (21 examples out of 51) of the cases where *kaela peal* (neck+on) is used as a complex unit (as in example (155)). There is one example where a plural PN is initially interpreted as being used with a freely combined phrase (see example (156)). However, on closer inspection it is decided that this example can also be analyzed as an in-between case. The TR *ike* is a polysemous word which may be interpreted in this context as referring to 'yoke' or to 'hardship', thus the phrase can be interpreted literally as well as non-literally, with the plural form of PN suggesting the latter interpretation.

---

[58] The data does not meet the requirements of the chi-square test. However, in this case, no such test is needed because the data is clearly interpretable.

(155) *Meie-ø   lapse-d   saa-vad   enda-ga   hakka-ma   ega*
we-GEN   child-PL   can-3PL   own-COM   cope with-SUP   nor
ela-ø   __meie-ø__   __kaela-ø__   __peal__ –   nii   nagu   meie   ei
live-CONNEG   we-GEN   neck.GEN   on.LOC   so   as   we   NEG
nõua-ø,   et   na-d   me-i-d   aita-ksi-d.
demand-CONNEG   that   they-PL   we-PL-PRT   help-COND-3PL
Lit. Our children can manage on their own and they are not living on our necks
and at the same time, we do not demand that they would help us.
'Our children can manage on their own; they are not dependent on us, and at the
same time, we do not demand that they would help us.' [naistekas.delfi.ee]

(156) *Ma   tõmba-si-n   ne-i-d   inimlik-e   sideme-te-ga,*
I   pull-PST-1SG   they-PL-PRT   human-PL.GEN   relation-PL-COM
armastuse-ø   pael-te-ga;   ma   ol-i-n   ne-i-le   nagu
love-GEN   ribbon-PL-COM   I   be-PST-1SG   they-PL-ALL   as
ikke-ø   kergitaja-ks   __nen-de__   __kaela-ø__   __pealt__;   ma
yoke-GEN   uplift-TRL   they-PL.GEN   neck-GEN   on.SEP   I
kummarda-si-n   nende   juurde   ja   toit-si-n   ne-i-d.
bow-PST-1SG   they-PL.GEN   towards   and   feed-PST-1SG   they-PL-PRT
Lit. I pulled them in with humane bonds, with the strings of love; it was as if I
was the uplifter of the yoke on their neck; I bowed down to them and fed them.
'I pulled them in with humane bonds, with the strings of love; it was as if I was
the uplifter of the yoke around their neck; I bowed down to them and fed them.'
[www.paevasona.ee]

Thus, it was observed that plural PNs are rather frequently used with *kaela peal*
(neck+on). The data showed that about one third of all the PNs of *kaela peal*
(neck+on) are in the plural form, and that over 40% of usages of *kaela peal*
(neck+on) as complex units in the plural form. These results are statistically
significant (*X*-squared = 9.2, *df* = 1, *p*-value = 0.002) with an intermediately
strong correlation (V = 0.35).

4.5.2.6. Summary and conclusions based on the distribution of singular and plural PNs

All of the studied phrases were used with singular as well as plural PNs. As
there is no set amount of plural PNs that could be considered a clear indication
of substantial decategorization, the the studied phrases can only be viewed
relative to each other. However, it can be assumed that the singular – as the
unmarked form – is always more frequent than the plural.

    The amount of plural PNs varies among the studied phrases from 3% to
31%. The data suggest that non-agreement is most common in case of *kaela
peal* (neck+on), with the amount of plural PNs accounting for 31% of the
examples. The plural PNs also appear in a considerable amount of examples in
cases of *käe all* (hand+under) (21%), *selja taga* (back+behind) (18%), and to a

lesser extent *käe kõrval* (hand+beside) (12%). Plural PNs are rarely used in the case of *külje all* (side+under) where they make up only 3% of the examples.

The distribution of singular and plural forms among the simple and complex structures shows that plural PNs occur more often with complex units in the case of all phrases except *külje all* (side+under). The fact that non-agreement tends to occur with the complex units is rather expected, because the use of plural PNs with a singular head noun would be semantically inappropriate in the context of body part terms (*mees-te selja taga* (*man*-PL.GEN back-PL.GEN)). Thus, in such cases the pluralization of both components would be expected (*mees-te selga-de taga* (man-PL.GEN back-PL.GEN behind) 'behind the men's backs'). However, in the case of complex units the use of the plural forms would not suggest semantic incompatibility, because in such cases, reanalysis has occurred and in this structure the PN is not modifying the bare noun but the whole complex unit. It also means that the phrase is lexicalised, and the association with the body part fading. Nevertheless, the fact that both of these structures are present in contemporary language and that the complex structures are still rather transparent, may also restrict the use of the plural in the case of complex units.

The largest amount (41%) of plural PNs occurred with the complex postpositional *kaela peal* (neck+on). The correlation between the plural PN and the complex structure was statistically significant and of intermediate strength. In the case of *selja taga* (back+behind) and *käe all* (hand+under), the plural forms were also quite frequent (25% and 22% respectively) and more inclined to be used with the complex unit; these results were statistically significant but the correlation was weak. The complex postpositional *käe kõrval* (hand+beside) was used with plural forms in only 11% of cases. However, the plural forms co-occurred with complex units and hybrid forms only. In the case of *külje all* (side+under), the plural forms were almost equally scarcely represented among the complex units and freely combined phrases, and were rather inclined to co-occur with freely combined phrases. However, in this case the correlation between the plural forms and the complex items is also weak. Thus, although the non-agreement seems to prefer the complex items, it is not (yet) a clear indicator of a complex unit status.

4.6. Productivity of the studied phrases

In the present chapter, the productivity of the studied phrases is discussed. Productivity is observed in the diversity of the sentential contexts of the studied phrases. I am primarily interested in the behavior of the phrases where they serve as complex adverbs and complex postpositions. The aim of the analysis is not to compare the diversity of the freely combined phrases and the complex units, but rather to observe the productivity of the complex structure. Thus, the following analysis focuses on the immediate sentential contexts of the complex structure only. As was suggested in section 2.5.4, the tendency of the studied

phrases to occur in certain fixed contexts is taken to suggest that the complex unit behaves as a lexical item; the tendency to occur with a larger array of PN and verb lemmas is taken to suggest grammatical behavior. It must be noted that the productivity is here described within the context of the studied phrases. That is, the terms 'productive' and 'non-productive' are used bearing in mind that (Estonian) adverbs and postpositions lay at the border of lexicon and grammar. Therefore, they cannot be expexted to be as productive as highly grammatical items, such as (grammatical) cases.

In the present study, the productivity is operationalized as the lexical diversity of two linguistic elements in the immediate proximity of the phrases – the lemmas of the PN and the verb. The PN whose semantic and grammatical properties were already discussed in sections 4.5.1 and 4.5.2 is the (pro)noun that precedes the body part related postpositional phrase and is part of the same phrasal structure (see section 1.3). As there might be several verbs within the same clause as the complex item, it must be noted that, here, we are primarily interested in the verb that together with the body part related phrase connects the LM and the TR (see section 2.5.4). It should be noted that the total number of examples where the PN lemma and the verb lemmas can be observed does not coincide. This is, due to the fact that PN lemmas can only observed in the examples where the complex structure behaves as a complex postposition. When it is realized as complex adverb, the PN is, of course, absent (see section 4.4). The verb lemma, on the other hand, may be observed in the case of complex adverbs as well as complex postpositions. However, in this case, the examples where the verb is elliptical are excluded.

The data was coded manually based on the principles described in section 4.5. In addition the following principles were implemented:

i   In the case of compound PNs, the PN lemma was coded as the last component of the compound if the compound was not lexicalized (*klaasi+kunstnik* 'glass artist' was coded as *kunstnik* 'artist'). Lexicalized compounds however, were coded as compounds (*osa+võtja* 'participant'). Proper nouns (e.g. anthroponyms, toponyms, company names) were coded as separate lemmas.

ii   Phrasal verbs were coded as separate lemmas only when they were lexicalized. For instance, *kaasa lööma* (lit. with + hit 'participate') was coded as a separate lemma, whereas *välja võtma* (lit. out take 'take out') was coded an instance of *võtma* 'take'.

iii The derivatives of the same lemma where coded as a single lemma, unless either of the derivatives were lexicalized. For instance, *jääma* 'remain' and *jätma* 'leave' were coded as the same lemma because they represent the same verb and, therefore, the same usage. However, *harjuma* 'accustom' and *harjutama* 'practice' were coded as separate lemmas because the meanings are rather distant and, as such, they represent different usages.

iv   In infinitive constructions, as in *läheb õppima* 'go-SG3 study-INF', only
      the infinitival verb (*õppima* 'study') is coded as it is more contentful than
      the finite verb, which is a quasi-auxiliary.

In the following, the analysis of PN and verb lemmas for each phrase are pre-
sented. Each section is structured as follows: first, the number of PN and verb
collocates are presented; second, the individual analyses of the strongest collo-
cates based on log-likelihood scores are presented; and finally, the amount of
productively combined PN lemmas and fixed expressions are presented.

4.6.1. *Käe all* (hand+under)

A close analysis of the immediate sentential context shows that as complex unit
(n=4054) *käe all* (hand+under) co-occurs with 1651 different PNs and 431
verbs. As the proper names were coded as separate lemmas, the vast amount of
proper names in the data results in the higher number of different PN lemmas.
When the examples with proper names are excluded from the dataset, the num-
ber of complex units is 2318 and the number of different PN lemmas is 368.
Quite expectedly, the vast majority of the PN and verb lemmas occur only a few
times. The data show that 96% of the PN lemmas[59] and 79% of the verb lemmas
occur up to five times. However, in both cases – the PN and the verb – there
is a small group of highly frequent lemmas.

4.6.1.1. The strongest collocates of *käe all* (hand+under)

As was mentioned in section 2.5.4, the frequency of the occurrence of indi-
vidual lemmas is also dependent on the overall frequency of these lemmas.
Thus, in order to determine the possible status of a fixed expression, it does not
suffice to merely observe the absolute frequency of the lemmas. In the
following, I present an analysis of the PN and verb lemmas that have the
strongest statistical association with *käe all* (hand+under) as a complex unit.
Table 17 shows the 20 most frequent PN lemmas of *käe all* (hand+under), their
frequency in etTenTen, the number of times they occur with *käe all*
(hand+under) as a complex unit (N collocate), and the log-likelihood score,
which shows the statistical strength between the PN lemma and the complex
unit. The lemmas are ranked according to the log-likelihood score, starting from
the highest value. The strongest collocates are divided with a dotted line.

---

[59]   When the examples with proper names were excluded, 86% of the lemmas occurred up
to five times.

Table 17. The strongest PN collocates of *käe all* (hand+under) as a complex unit

PN lemma	meaning	*n* lemma (in etTenTen)	*n* (collocate)	log-likelihood score
juhendaja	supervisor	9,499	177	2,524
treener	coach	18,126	158	2,013
õpetaja	teacher	125,700	174	1,577
meister	master	17,287	122	1,503
tema	s/he	1,765,208	306	1,501
kes	who	1,036,055	218	1,154
spetsialist	specialist	29,060	53	509
instructor	instructor	1,826	25	341
professionaal	professional	3,994	26	316
mentor	mentor	2,805	22	276
dirigent	conductor	4,991	22	250
lavastaja	director	6,574	22	238
kokk	chef	7,620	21	219
õppejõud	lecturer	15,357	22	201
koolitaja	leader	7,488	19	195
nemad	they	178,486	37	195
juht	educator	257,232	41	194
isa	father	178,746	35	180
proff	pro	2,479	15	180
asjatundja	expert	8,165	18	180

Table 17 shows that the strongest PN collocates of the complex unit *käe all* (hand+under) are the following – *juhendaja* 'supervisor', *treener* 'coach', *õpetaja* 'teacher', *meister* 'master', *tema* 's/he', and *kes* 'who'. As explained in section 2.5.4 the higher values suggest stronger associations and the lower values weaker association between the PN and the phrase. As the log-likelihood score is also calculated based on the individual frequencies of the co-occurring elements, the ranking of the lemmas by the log-likelihood score does not co-incide exactly with the ranking based on the number of occurrences (*n* collocate). For instance, the most frequent lemmas *tema* 's/he' and *kes* 'who' based on absolute frequency are in 5th and 6th position in the table, because these lemmas are more frequent in general. However, the six most frequent lemmas are also most strongly associated with *käe all* (hand+under) based on the log-likelihood score.

The strongest collocates are considered to form more or less fixed expressions (e.g. *juhendaja käe all*, *meistri käe all*). The status of fixed expressions is taken to be indicated by the large differences in the log-likelihood scores. Although there is no certain log-likelihood value that determines the status of a fixed expression or a freely combined phrase, Table 17 shows that

the scores of the top six lemmas are considerably higher than the other lemmas. Among the strongest collocates, two lemmas – *juhendaja* 'supervisor' and *treener* 'coach' – stand out as they have by far the highest log-likelihood scores (2,524 and 2,013 respectively). These lemmas are followed by *õpetaja* 'teacher', *meister* ' master', and *tema* 's/he' with log-likelihood scores around 1500, and finally *kes* 'who', which has the weakest association (1154) among the top six lemmas, but is still considered to belong to the group of strong collocates of *käe all* (hand+under). The rest of the lemmas presented in the table, starting with *spetsialist* 'specialist' have considerably lower log-likelihood scores (≤509), and are not considered to be strong collocates of *käe all* (hand+under). This is also supported by the absolute frequencies. Figure 14 shows that there is a large gap between the number of occurrences of the most frequent and less frequent lemmas.

Even though the strong collocates of *käe all* (hand+under) are considered as somewhat fixed expressions, they are (in contemporary language) not idiomatic, i.e. *käe all* (hand+under) does not form a larger lexicalized item with its strong collocates. For instance, *juhendaja käe all* (supervisor-GEN hand+under; 'under a supervisor') (see example (157)) and *treeneri käe all* (coach-GEN hand+under; 'under a coach') (see example (158)) do not express a meaning that is clearly distinct from that of *asjatundja käe all* (159) or *profi käe all* (160). The same applies for the other frequent collocates – *treener* 'coach', *õpetaja* 'teacher', *meister* 'master', *tema* 's/he', and *kes* 'who'. Therefore, these usages are considered as more typical examples of the complex unit *käe all* (hand+under).

(157) *Järgne-si-d* *õpiaasta-d* *erineva-te* **_juhendaja-te_**
follow-PST-3PL apprenticeship-PL different-PL.GEN supervisor-PL.GEN
käe-ø   **_all_**   *Aulnay-sous-Bois'-s*.
hand-GEN under.LOC Aulnay-sous-Bois'-INE
Lit. This was followed by an apprenticeship under the hand of various
supervisors in Aulnay-sous-Bois.
'This was followed by an apprenticeship under the supervision of various tutors
in Aulnay-sous-Bois.' [www.kitarrifestival.ee]

(158) *Investeeri-da* *tule-b* *noor-te-sse* *ja* *par-ima-d* *tule-b*
invest-INF must-3SG youngster-PL-ILL and good-SPL-PL must-3SG
koonda-da *par-ima-te* **_treeneri-te_** **_käe-ø_** **_alla_**.
bunch-INF good-SPL-PL.GEN trainer-PL.GEN hand-GEN under.LAT
Lit. Young people should be invested into and the best should be gathered under
the hand of the best trainers.
'Young people should be invested into and the best should be gathered under the
best trainers.' [www.epl.ee]

(159) *Lähe-n   varsti   **asjatundja-te   käe-ø   alla   õppi-ma***
go-1SG   soon   expert-PL.GEN   hand-GEN   under.LAT   study-SUP
ja   siis   jaga-n   kogemus-i,   kuidas   ise
and   then   share-1SG   experience-PL.PRT   how   self
enda-le   looduslik-e-st   koostisaine-te-st   kreem-e   valmista-da.
own-ALL   natural-PL-ELA   ingredient-PL-ELA   cream-PL.PRT   make-INF
Lit. Soon enough I will be studying under the hand of experts and then I will share my experience on how to make creams out of natural ingredients.
'Soon enough I will be studying under experts and then I will share my experience on how to make creams out of natural ingredients.' [www.bioneer.ee]

(160) ***Profi-ø   käe-ø   all   omanda-tud   teooria-t***
professional-GEN   hand-GEN   under.LOC   acquire-PST.PTCP   theory-PRT
ja   praktika-t   kontrolli-t-i   Junori-ø   juuksuri-te
and   practice-PRT   control-IMPS-PST   Junor-GEN   hairdresser-PL.GEN
õppeklassi-ø   eksami-te-l.
learning class-GEN   exam-PL-ADE
Lit. The theories and practice acquired working under professionals were put to the test during exams in the Junior hairdressers' learning class.
'The theories and practice acquired working under professionals were put to the test during exams in the Junior hairdressers' learning class.' [www.iluguru.ee]

Moreover, Table 17 shows that the strongest collocates of the complex *käe all* (hand+under) are semantically uniform, i.e. all refer to an authority figure – people that are superior because of their knowledge, skills and or their position or role. This also applies to the frequent collocate *kes* 'who', which, as a pronoun, has a very general meaning. In the data, it mostly (see example (161)) occurs as an interrogative relative pronoun the correlate of which in the main clause refers to an authority figure (here: *tippjuht* 'top manager').

(161) *Aleksandras   ehk   Alex,   nagu   ta   en-d   kutsu-da*
Aleksandras   also known as   Alex   as   s/he   own-PRT   call-INF
*eelista-b,   on   **tippjuht,   kelle-ø   käe-ø   all***
prefer-3SG   be.3SG   top manager   who-GEN   hand-GEN   under.LOC
on   Delfi   kasva-nud   Leedu-ø   suur-ima-ks
be.3SG   Delfi   grow-PST.PTCP   Leedu-GEN   big-SPL-TRL
uudisteportaali-ks. [www.epl.ee]
news portal-TRL
Lit. Aleksandras, also known as Alex, is an exceeutive manager under whose hand Delfi has grown to be Lithuania´s biggest news portal.
'Aleksandras, also known as Alex, is an exceeutive manager; under his rule Delfi has become Lithuania´s biggest news portal.' [www.epl.ee]

This suggests, that the use of *käe all* (hand+under) is still thematically restricted. Even though it co-occurs with a variety of different lemmas, it still prefers a narrow semantic class. However, such restriction seems inevitable

given the meaning of the complex unit. A function word that expresses the meaning 'under mental/physical control' is bound to keep the company of words that refer to authority figures. The prevailing of such lemmas is also connected to the fact that the usages where *käe all* (hand+under) expresses mental control are much more frequent than physical control (see section 4.3.1), because lemmas that refer to authority figures are more inclined to occur with this sense. A similar phenomenon has also been observed by Hoffmann (2005: 79) in the complex expression *by virtue of*, which he found to frequently collocate with nouns belong to legal jargon.

I now turn to discuss the individual verb lemmas that occur with *käe all* (hand+under) focusing on the most frequent verbs. Table 18 gives 20 strongest verbal collocates of *käe all* (hand+under), their frequency in etTenTen, the number of times they occur with *käe all* (hand+under) (N collocate), and the log-likelihood score, which shows the statistical strength between the verb lemma and *käe all* (hand+under). The lemmas are ranked according to the log-likelihood score. The strongest collocates have been separated with the dotted line.

Table 18. The strongest verbal collocates of *käe all* (hand+under) as a complex unit

verb lemma	meaning	*n* lemma in etTenTen	*n* Collocate	log-likelihood score
õppima	study	117,240	640	7,547
valmima/valmistama	be made/make	89,662	376	4,235
treenima	train	9,090	134	1,846
harjutama	practice	12,503	107	1,358
alustama	begin	105,215	93	758
saama	get	1,871,918	192	738
töötama	work	137,400	90	680
mängima	play	91,053	80	651
omandama	acquire	28,075	61	607
kasvama	grow	107,601	68	509
täiendama	better	19,380	48	490
tulema	come	960,238	114	471
läbi käima	go through	14,975	44	464
tegutsema	take action	55,913	51	419
sündima	be born	62,712	51	407
jätkama/jätkuma	continue/be continued	103,425	55	392
laulma	sing	25,537	36	327
tegema	do	1,091,109	92	318
sirguma	thrive	2,157	23	302
arenema/arendama	develop/be developed	72,619	39	279

The data suggest that *õppima* 'study' is by far the strongest verbal collocate of *käe all* (hand+under) with the log-likelihood score of 7547, *valmima/valmistama* is a close second with a log likelihood value of 4235. *Treenima* 'train' and *harjutama* 'practice' may also be considered as strong collocates because they have two times higher log-likelihood values (1846 and 1358 respectively) than the following lemmas in the table (log-likelihood score 758 and below) and they were also frequent collocates based on their absolute frequency of co-occurrence with *käe all* (hand+under) (134 and 107 instances respectively).

It was discussed above that the complex unit *käe all* (hand+under) tends to take PNs that refer to authority figures, such as *juhendaja* 'supervisor', *treener* 'trainer', *õpetaja* 'teacher', etc. Accordingly, the strongest verbal collocates of *käe all* (hand+under) express the activities that can be done under the supervision of such figures. For example, *õppima* 'study' (162), *treenima* 'train' and *harjutama* 'practice' (163). Among the 20 strongest collocates (in Table 18) there are many verbs that refer to such activities, e.g. *töötama* 'work', *mängima* 'play (sports)', *omandama* 'acquire', *täiendama* 'better', *tegutsema* 'take action', etc. Although such correspondence is quite expected, it is not the case that the strongest verbal and nominal collocates always co-occur. For instance, in the total of 640 instances, *õppima* 'study' co-occurs with 350 different PN lemmas (99, if the proper names are excluded) and *valmima*/*valmistama* 'be made/make' with 214 different PN lemmas (69, if the proper names are excluded). Nevertheless, the association between these elements has not been measured.

(162) *Lisaeriala-na   sa-i   noor   Külli Merike   Aarma-ø*
       minor-ESS   get-PST.3SG   young   Külli Merike   Aarma-GEN
       käe-ø   all   ka   koorijuhtimis-t   õppi-da.
       hand-GEN   under.LOC   also   conducting-PRT   study-INF
       Lit. As a minor, Külli also got to learn conducting under the hand of Merike Aarma.
       'As a minor, Külli also got to learn conducting under the supervision of Merike Aarma.' [www.lounaleht.ee]

(163) *Ridala-s   ela-va-d   poisi-d   on   **harjuta-nud***
       Ridala-INE   live-PTCP-PL   boy-PL   be.3PL practice-PST.PTCP
       *lapsevanema-st ringijuhendaja Aive Aljaste-ø   **käe-ø   all.***
       parent-ELA   instructor   Aive Aljaste-GEN   hand-GEN   under.LOC
       Lit. The boys who live in Ridala have been practicing under the hand of Aive Aljaste, one of the parents who is also their instructor.
       'The boys from Ridala have been practicing under the supervision of Aive Aljaste, one of the parents who is also their instructor.' [www.ridala.edu.ee]

The verb *valmima*/*valmistama* 'be made / make' is most often used in examples where *käe all* (hand+under) is used to refer to a concealed agent (164) and (165). In (164) *käe all* (hand+under) refers to PHYSICAL CONTROL but the verb may also occur in examples where *käe all* (hand+under) refers to MENTAL CONTROL (165).

(164) *Sama-s   kompleksi-s   asu-b   ka   meie-Ø   kooli-Ø   õmbleja,*
Same-INE   complex-INE   lie-3SG   also   we-GEN   school-GEN   tailor
*kelle-Ø   **käe-Ø   all**   **valmi-vad**   Capeti-Ø*
who-GEN   hand-GEN   under.LOC   be made-3PL   Capeti-GEN
kaubamärki-Ø   kand-va-d   riide-d   võistlus-te-ks
trade mark-PRT   carry-PTCP-PL   cloth-PL   competition-PL-TRL
ja   treeningu-te-ks.
and   training-PL-TRL
Lit. In the same complex our school tailor is situated under whose hand Capeti trademark clothes for completion as well as training are made.
'The Capeti trademark clothes used for completion as well as training are made by our school tailor who works in the same complex.' [www.laguun.ee]

(165) *Se-da,   et   film   on   amatööri-de   **käe-ø   all***
this-PRT   that   film   be.3SG   amateur-PL.GEN   hand-GEN   under.LOC
valmi-nud,   *või-b   näh-a   ka   inimes-te*
be made-PST.PTCP   can-3SG   see-INF   also   people-PL.GEN
näitlemisoskuse-st.
acting skill-ELA
Lit. That the film has become ready under the hand of amateurs can also be seen from the people´s acting skill.
'That the film has been done by amateurs can also be seen from the people´s acting skill.' [filmitalgud.ee]

Similarly to the results of the analysis of the PN lemmas, the data suggests that *käe all* (hand+under) does not form idiomatic expressions with its strongest verbal collocates either. For instance in (162) *käe all* (hand+under) does not carry a meaning that is any different than in (165). Thus, in this case, too, the strong collocates are considered to represent just typical examples of *käe all* (hand+under) as a complex unit. However, due to their high frequency and strong association with *käe all* (hand+under) as a complex unit, examples with these lemmas are not considered to indicate productive use of the complex unit because such usages are more or less fixed. Thus, in the following, the productivity of *käe all* (hand+under) is observed based on the amount of usages with strong and weak collocates.

4.6.1.2. The productive use of *käe all* (hand+under) as a complex unit

It can be observed in tables 17 and 18 that the strong PN and verbal collocates of *käe all* (hand+under) are highly frequent and, thus, it can be expected that the usages with these lemmas make up a considerable amount of the available examples. As fixed expressions, such examples would not suggest a productive use of *käe all* (hand+under) as a complex unit. Thus, in order to observe the amount of productively formed examples in the dataset, one needs to determine the proportion of such usages among all of the examples of *käe all* (hand+under) as a complex structure.

Figure 14 shows the cumulative percentage of the instances of *käe all* (hand+under) as a complex unit based on the frequency of the PN lemma. The horizontal axis gives the number of the occurrences of the lemmas (lemmas that occur 1–5 times, 6–10 times, 11–20 times, etc.), and the left vertical axis gives the absolute number of such examples; the right vertical axis gives the percentage of the examples formed with each number class of lemmas in the dataset.
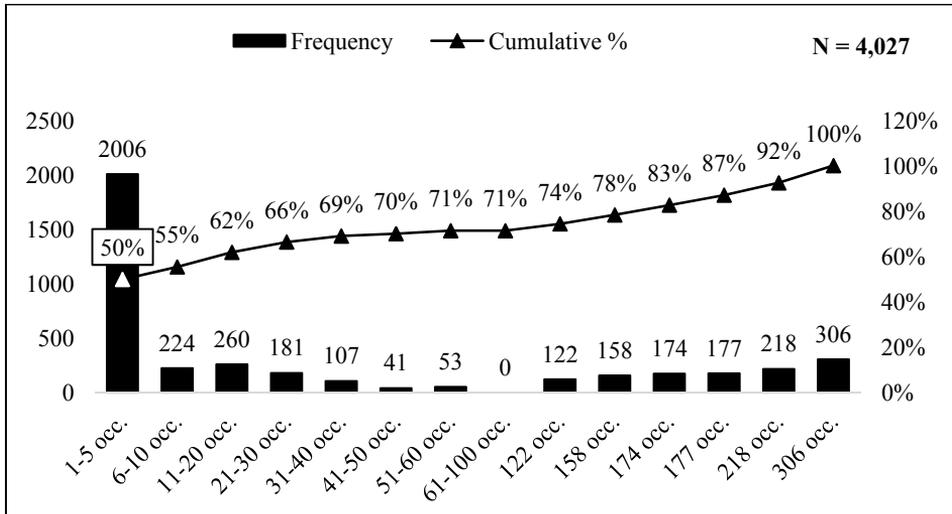

Figure 14. The cumulative percentage of examples of *käe all* (hand+under) formed with PN lemmas based on the number of occurrence

It can be observed in Figure 14 that 50% of the data (2006 examples out of 4027) are formed with lemmas that occur only 1–5 times, 55% of the examples (2230 out of 4027) are formed with lemmas that occur up to 10 times and 62% of the examples (2490 out of 4027) are formed with lemmas that occur up to 20 times. Based on the association measures presented in Table 17, the lemmas that occurred up to 53 times are not considered strong collocates and, therefore, such examples are considered to be formed productively. As such examples make up 71% (2872 out of 4027) of the examples, the use of the complex unit *käe all* (hand+under) can be considered quite productive.

Quite expectedly, the vast number of proper names among the PNs of *käe all* (hand+under) affects the proportion of examples where *käe all* (hand+under) is used productively. When the examples that are formed with proper names are excluded from the data set, the uses with the 6 most frequent collocates make up about half (49.1%, 1138 examples out of 2318) of all the usages of *käe all* (hand+under) as a complex item. Thus, the abundant use with proper names contributes to the productivity of *käe all* (hand+under).

Figure 15 gives the cumulative percentage of the instances of *käe all* (hand+under) as complex unit based on the frequency of the verb lemma.
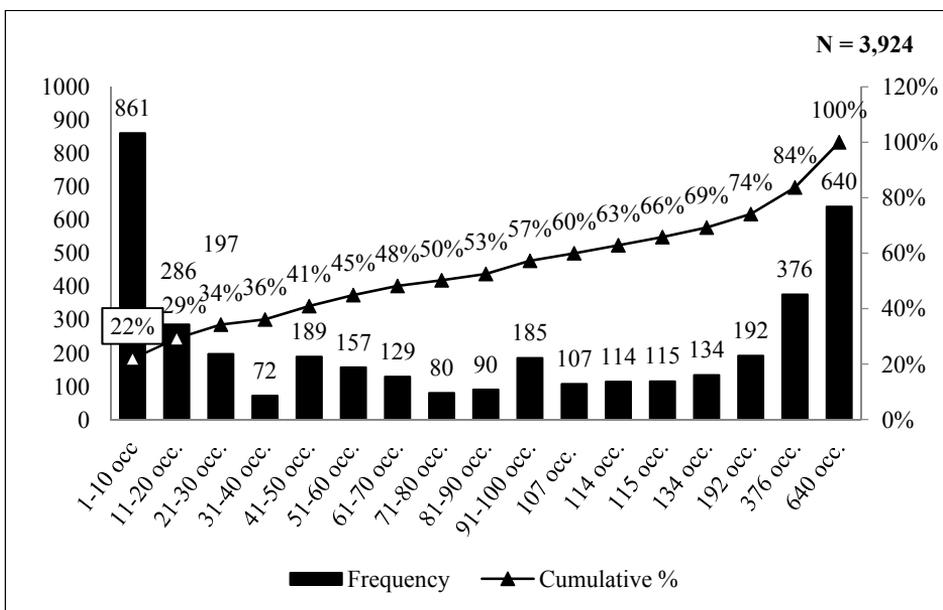
Figure 15. The cumulative percentage of examples of *käe all* (hand+under) formed with verb lemmas based on the number of occurrence

It can be observed in Figure 15 that two of the most frequent verb lemmas – *õppima* 'study' and *valmima*/*valmistama* 'be made/make' – that were also the strongest collocates of *käe all* (hand+under) are used in 26% of the examples where *käe all* (hand+under) is used as complex unit. Because the lemmas *treenima* 'train' and *harjutama* 'practice' which occurred on 134 and 107 occasions respectively were are also considered as strong collocates of *käe all* (hand+under), they belong with the unproductively formed usages. Therefore, the total number of examples formed with the strong collocates is 1185 (out of 3924), which makes up 32% of the examples where *käe all* (hand+under) behaves as a complex unit. Thus, the remaining 68% of the examples are formed productively, which is indicative of high productivity of *käe all* (hand+under) as a complex unit.

4.6.1.3. Summary of the productivity of *käe all* (hand+under)

In conclusion, based on the analysis of PN and verb lemmas, the use of *käe all* (hand+under) as a complex unit is quite productive. In 2047 examples, the complex unit *käe all* (hand+under) occurs with 1651 PN lemmas and 431 verb lemmas. Most of the lemmas are low frequency collocates – 96% of the PN lemmas and 79% of the verb lemmas occur only up to 5 times – , and only small amount of lemmas co-occur frequently with *käe all* (hand+under).

It was observed that most of the frequent PN lemmas refer to an authority figure. This is associated with the semantics of the complex unit. As a complex

unit *käe all* (hand+under) mostly refers to MENTAL CONTROL that the LM has over the TR. The examples with the strongest PN collocates – *juhendaja* 'supervisor', *treener* 'coach', *õpetaja* 'teacher', *meister* 'master', *tema* 's/he, and *kes* 'who' – make up 29% of the data. Thus, based on the analysis of PN lemmas, 71% of the examples can be considered to be formed productively. However, it must be noted the high productivity of *käe all* (hand+under) is dependent on its abundant use with proper names. Excluding the examples with proper names resulted in just about half of the examples being formed productively.

Based on the analysis of verb lemmas, five frequent collocates that have a strong statistical association with *käe all* (hand+under) were determined. It was found that in accordance with the semantic unity of the PN lemmas, these verbs also express activities that can be performed under the supervision of authority figures – *õppima* 'study', *valmima*/*vamistama* 'be made / make', *treenima* 'train', and *harjutama* 'practice'. Based on the analysis of verb lemmas, 68% of the examples were formed productively. Thus, despite of some thematic inclinations in the PN and verb lemmas, *käe all* (hand+under) as a complex unit is considered to be quite productive.

4.6.2. *Külje all* (side+under)

The phrase *külje all* (side+under) occurs as a complex unit in 2530 instances. Within these 2530 examples, *külje all* (side+under) occurs with 591 PN lemmas and 333 verb lemmas. In this case, too, the data show that 90% of PN lemmas and 84% of verb lemmas occur up to five times and that there is only a small amount of the lemmas are highly frequent. In the following section, these frequent lemmas are analyzed in order to determine the strongest PN and verbal collocates of *külje all* (side+under).

Similarly to the phrase *käe all* (hand+under) discussed above in section 4.6.1, *külje all* (side+under) also frequently (78% of all lemmas) co-occurs with proper names. As stated previously, this increases the number of low-frequency lemmas. In the case of *külje all* (side+under), the proper names are predominantly toponyms. When these examples are excluded, 87% of the lemmas (114 out of 131) occur up to five times. Thus, the diversity of the PNs of *külje all* (side+under) is not dependent on its abundant use with toponyms.

4.6.2.1. The strongest collocates of *külje all* (side+under)

In order to determine the amount of productively formed examples, one must first determine the fixed expressions. Table 19 shows the most frequent PN collocates of *külje all*, their frequency in the etTenTen corpus, the number of times they occur with *külje all* (side+under) as a complex postposition, and their log-likelihood score. The lemmas are ranked based on the log-likelihood score. The strongest collocates have been divided with a dotted line.

Table 19. The strongest PN collocates of *külje all* (side+under) as a complex unit

PN lemma	meaning	*n* lemma (in etTenTen)	*n* (collocate)	log-likelihood score
Tallinn	Tallinn*	227,949	334	3,450
linn	city/town	155,973	155	1,481
Tartu	Tartu*	153,133	142	1,337
Põlva	Põlva*	12,485	48	588
Rakvere	Rakvere*	19,654	48	545
Võru	Võru*	32,023	51	535
Pärnu	Pärnu*	57,237	55	522
Haapsalu	Haapsalu*	21,541	38	407
Otepää	Otepää*	10,623	32	377
Viljandi	Viljandi*	28,035	33	326
Elva	Elva*	5,989	26	325
Põltsamaa	Põltsamaa*	5,186	23	288
Narva	Narva*	30,223	28	264
Helsingi	Helsinki*	12,837	22	234
vanalinn	old town	10,897	19	203
Jõgeva	Jõgeva*	7,593	17	190
alevik	hamlet	4,655	15	179
Venemaa	Russia*	98,260	24	162
kesklinn	centre	17,704	17	161
Rapla	Rapla*	8,727	15	160

* Toponym

Table 19 shows that the strongest PN collocates of *külje all* (side+under) are *Tallinn* (334 occurrences), *linn* 'city/town' (155 occurrences), and *Tartu* (142 occurrences). The log-likelihood scores of these lemmas are 3,450, 1,481, and 1,337 respectively, which are considerably higher than that of the sub-sequent lemmas in the table. For example the next highest instance, the place name *Põlva* – which co-occurred with *külje all* (side+under) in 48 instances – has a log-likelihood score of 588. Thus, it seems that the utterances *Tallinna külje all* (Tallinn-GEN side+under; 'close to Tallinn')*, linna külje all* (city/town-GEN side+under; 'close to city/town')*,* and *Tartu külje all* (Tartu-GEN side+under) are most strongly fixed usages among the examples of *külje all* (side+under) as a complex postposition. However, similarly to *käe all* (hand+under), the combination of *külje all* (side+under) and its most frequent PNs do not form holistic expressions that would carry a distinct phrasal meaning. On the contrary, in combination with *külje all* (side+under) the most frequent lemmas express a similar meaning to all of the other examples in Table 19. For instance, *Tallinna külje all* 'close to Tallinn' (see example (166)) is not semantically much dif-

ferent than the utterance *Rapla külje all* 'close to Rapla' (see example (167)). The latter only occurs in 15 instances and has a log-likelihood score of 160, which suggests a much weaker statistical link.

(166)   *...ei   usu   et   tartlas-te-l   nii   palju*
       NEG   believe-CONNEG   that   Tartu resident-PL-ADE   so   much
       *raha-ø   on   et   bussi-ø   või   auto-ga   **tallinna-ø***
       money-PRT   *be.3SG   that*   *bus*-GEN   or   car-COM   Tallinn-GEN
       külje   al   sõit-a?!*
       side-GEN   under.LAT   drive-INF
       Lit. ... don't believe that residents of Tartu have enough money to take a bus or drive a car under the side of Tallinn?
       '... don't believe that residents of Tartu have enough money to take a bus or drive a car near Tallinn?' [rahvahaal.delfi.ee]

(167)   *Täna   ava-b   AS   Ingle   **Rapla-ø**   **külje-ø**   **all***
       today   open-3SG   PLC   Ingle   Rapla-GEN   side-GEN   under.LOC
       Eesti-ø   esimese-ø   euronõue-te-le   vastava-ø
       Estonia-GEN   first-GEN   EU requirement-PL-ALL   fit-GEN
       vedelkemikaali-de   terminali-ø.
       liquid chemical-PL.GEN   terminal-GEN
       Lit. Today, Ltd. Ingle opens Estonia's first terminal of liquid chemicals that meets the EU requirements, under the side of Rapla.
       'Today, Ltd. Ingle opens Estonia's first terminal of liquid chemicals that meets the EU requirements, near Rapla.' [keemia.ee]

The semantic closeness between the utterances formed with the more frequent and less frequent lemmas is connected with the fact that both of these lemmas are toponyms. As mentioned previously, *külje all* (side+under) frequently collocates with toponyms in general. Moreover, even the regular nouns that co-occur with the complex postpositional *külje all* (side+under) mostly refer to (geographical) locations. This is in line with the observation made in section 4.5.1.2 that most often (92%) the PN refers to a region. Thus, it seems that the usages where *külje all* (side+under) refers to physical proximity to a region are rather uniform.

    However, the question remains, why are *Tallinn* and *Tartu* most frequent among the toponyms? It is likely associated with the fact that Tallinn and Tartu are the two largest cities in Estonia. Tallinn, which is situated on the northern coast of Estonia, can be considered as the regional capital of northern Estonia, and Tartu, which is situated in the south, serves as the regional capital of southern Estonia. Thus, it is likely that these two toponyms are used as landmarks within their respective regions. That is in northern Estonia, the locations of other places and entities (especially the ones that are close by) are profiled in relation to Tallinn, and in southern Estonia in relation to Tartu. Perhaps even more importantly, due to urbanization the largest cities are expanding creating new municipalities in close proximity of these cities and the speakers' need to refer to them.

I now turn to discuss the individual verb lemmas that occur with *külje all* (side+under), focusing on the strongest collocates. The results in Table 20 show that *asuma* 'lie' is by far the strongest verbal collocate of *külje all* (side+under) with the log-likelihood score of 5468. Even though the log-likelihood score of *pugema* 'creep' is considerably lower (1138) than in case of *asuma* 'lie', the latter it is still considered to be strongly associated with *külje all* (side+under) because these words form an idiomatic expression (see the discussion below). The log-likelihood scores of the rest of the lemmas are considerably lower, and as such these lemmas do not seem to belong to the group of strong collocates of *külje all* (side+under).

Table 20. The strongest verbal collocates of *külje all* (side+under) as a complex unit

verb lemma	meaning	*n* lemma in etTenTen	*n* collocate	log-likelihood score
asuma	lie	118,320	448	5,468
pugema	creep	6,241	78	1,138
elama	live	201,395	112	937
olema	be	9559,606	322	889
ujuma	swim	10,425	53	678
rajama	found	30,566	58	628
toimuma	happen	252,280	82	598
paiknema	be located	12,846	35	404
tegutsema	take action	55,912	36	312
avama	open	102,004	36	269
kolima	move	19,550	20	192
ronima	climb	15,605	19	189
ehitama	build	74,919	23	165
jõudma	reach	233,096	28	149
sõitma	drive	130,035	24	148
kerkima	arise	16,695	12	107
trügima	force one's way	3,134	9	105
parkima	park	7,093	10	102
sündima	be born	62,712	15	100
ostma	buy	134,365	18	99
leidma/leiduma	find / be found	459,062	26	99

Similarly to the analysis of *käe all* (hand+under) presented above (in section 4.6.1.1), the verbal collocates are also semantically in correspondence with the strongest nominal collocates. It was demonstrated above that *külje all* (side+under) tends to co-occur with PNs that refer to locations – the 20 strongest collocates of *külje all* (side+under) were mostly toponyms. Accordingly, the most

typical verbal collocate of *külje all* (side+under) refer to being situated relative to the area expressed by the PN (as in (168)). Additionally, there are more verbs of location among the 20 strongest collocates presented in Table 20 (e.g. *olema* 'be', *paiknema* 'be located' (as in (169)). However, there are also verbs of other semantic classes, e.g. action verbs (as *tegutsema* 'take action' in (170)).

(168) *Eestlane-gi   ei   viitsi-nud   järjekorda-de   pärast*
Estonian-CL   NEG   bother-PST.PTCP   queue-PL.GEN   because of
*avasta-da   meie-ø   **külje-ø**   **all**   **asu-va-t***
discover-INF   we-GEN   side-GEN   under.LOC   lie-PTCP-PRT
Peterburi-ø   ja   Pihkva-t.
Petersburg-PRT   and   Pskov-PRT
Lit. Even Estonians could not be bothered to discover St. Petersburg and Pskov right under our side because of the queues.
'Even Estonians could not be bothered to discover St. Petersburg and Pskov right next to us because of the queues.' [arvamus.postimees.ee]

(169) *Ainuüksi   Tartu-ø   **külje-ø**   **all**   **paikne-va***
only   Tartu-GEN   side-GEN   under.LOC   be located-PTCP.GEN
Saareki-ø   aastane   vajadus   on   3000 tonni-ø.
Saarek-GEN   annual   need   be.3SG   3000 ton-PRT
Lit. Merely the annual need of Saareki, situated under the side of Tartu, is more than 3000 tons.
'Merely the annual need of Saareki, situated near Tartu, is more than 3000 tons.' [www.maaleht.ee]

(170) *Reola-ø   Kartuliühistu-le   pan-i-d   aluse-ø*
Reola-GEN   Potato-co-operative-ALL   put-PST-3PL   basis-GEN
*Tartu-ø   **külje-ø**   **all**   **tegutse-va-d***   *Ülenurme*
Tartu-GEN   side-GEN   under.LOC   take action-PTCP-PL   Ülenurme
valla-ø   kartulikasvataja-d.
parish-GEN   potato grower-PL
Lit. Reola Potato-co-operative was set up by potato growers of Ülenurme parish, under the side of Tartu.
'Reola Potato-co-operative was set up by potato growers of Ülenurme parish, near Tartu.' [paber.maaleht.ee]

The verb *pugema* 'creep' is also considered to be a strong collocate of *külje all* (side+under). Despite of its considerably lower log-likelihood score, usages as exemplified in (171) are considered to be fixed expressions because they carry a holistic meaning 'to get in with someone' and only occur in the lative form. Even though such usages are semantically rather close to the other uses that express the function MENTAL PROXIMITY (see section 4.3.2), they cannot be considered be formed productively.

(171) *Juba   on   na-d   asjatundja-te-na   ka   uue-ø*
      already   be.3PL   they-PL   expert-PL-ESS   also   new-GEN
      *rahvastikuministri-ø   **külje-ø   alla   puge-nud**   ja*
      population minister-GEN   side-GEN   under.LAT   creep-PST.PTCP   and
      paku-vad   lahenduse-na   välja   eestlas-te   integreerimis-t.
      offer-3PL   solution-ESS   out   Estonian-PL.GEN integration-PRT
      Lit. Already they have creeped under the side of the new Minister of Population
      and are offering the integration of Estonians as a solution.
      'Already they have cozyed up to the new Minister of Population and are offering
      the integration of Estonians as a solution.' [www.syndikaat.ee]

4.6.2.2. The productive use of *külje all* (side+under) as a complex unit

In the case of *külje all* (side+under), the strongest collocates are not considered
to serve as examples of productive use of the complex functional word either.
Thus, it is also useful to determine the amount of such examples. Figure 16
shows the cumulative percentage of the instances of *külje all* (side+under) as a
complex postposition based on the frequency of the PN lemma. The horizontal
axis gives the number of occurrences of the lemmas; and the left vertical axis
gives the absolute number of such examples; the right vertical axis gives the
percentage of the examples formed with each number class of lemmas.
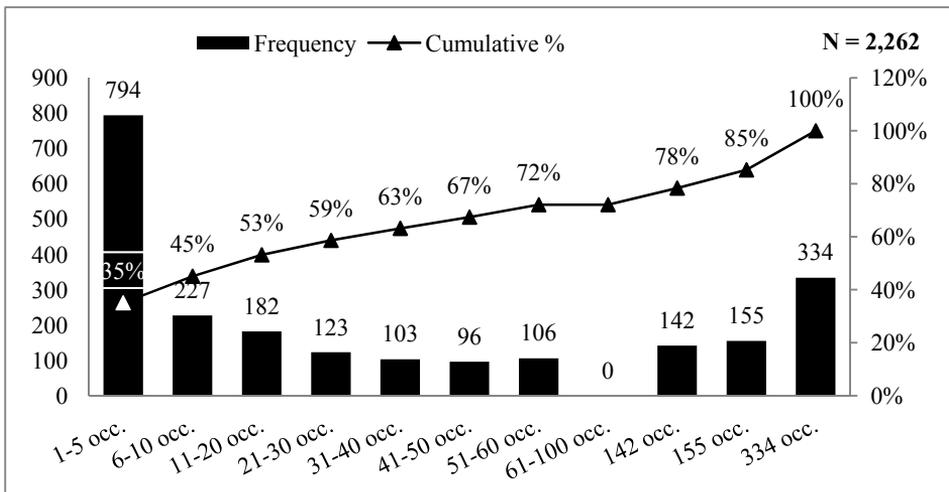
Figure 16. The cumulative percentage of examples of *külje all* (side+under) formed
with PN lemmas based on the number of occurrence

Figure 16 shows that although only 35% of the examples (749 out of 2262) are
formed with PN lemmas that occur up to 5 times, 45% of the examples (1021
out of 2262) are formed with lemmas that occur up to 10 times, and 53% of the
examples are formed with lemmas (1203 out of 2262) that occur up to 20 times.
Based on the association measures presented in Table 19, the lemmas that

occurred on up to 55 occasions are not considered strong collocates and are thus considered to account for the productive use of the complex postposition. As such examples make up 72% of the data (1631 out of 2262), the use of the complex postpositional *külje all* (side+under) is considered quite productive.

As shown in Table 19, most of the frequent PN lemmas are toponyms. Thus, it makes sense to also view the distribution of more frequent and less frequent lemmas in the sample once the proper names have been excluded (*n* = 498). In this case, there is only one lemma that stands out as highly frequent (*linn* 'town/city' occurs on 155 occasions), the rest of the PNs only occur up to 20 times. The data shows that almost half (49%) of the examples of *külje all* (side+under) used as a postposition are formed with lemmas that appear 1–10 times, and 69% with lemmas that appear up to 20 times. The highly frequent *linn* 'city/town' occurs in 31% of the examples. This is about the same amount that the most frequent lemmas make up among all of the examples of *külje all* (side+under) as a complex postposition. Thus, the productive use of *külje all* (side+under) as a complex unit is not dependent on its abundant use with toponyms.

It can be observed in Figure 17 that the examples with strongest verbal collocates – *asuma* 'lie' and *pugema* 'creep' – are used in 34% of the examples where *külje all* (side+under) is used as complex unit. Thus, the remaining 66% of the examples are formed productively.
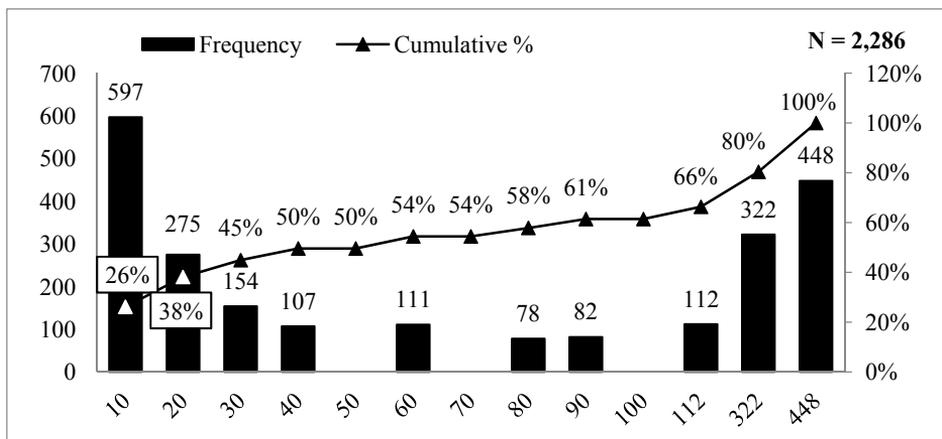
Figure 17. The cumulative percentage of examples of *külje all* (side+under) formed with verb lemmas based on the number of occurrence

4.6.2.3. Summary of the productivity of *külje all* (side+under) as a complex unit

In conclusion, the analysis of PN and verb lemmas suggests that the use of the complex unit *külje all* (side+under) is quite productive. In 2530 examples, the complex unit *külje all* (hand+under) occurs with 591 PN and 333 verb lemmas.

The data suggest that most of the lemmas are low frequency collocates – 90% of the PN lemmas and 84% of the verb lemmas occur up to 5 times. Only a few lemmas are highly frequent collocates of *külje all* (side+under).

The PNs of *külje all* (side+under) typically refer to places, which are often (78%) referred to with toponyms. The frequent PN lemmas that have a strong statistical association with *külje all* (side+under) are the toponyms *Tallinn* and *Tartu*, and the nominal linn 'town/city'. The examples formed with these strong collocates make up about a third (28%) of all examples. Thus, the rest of the 72% of the examples are formed productively.

As for verbs, there is one extremely frequent collocate that also has a strong statistical association with *külje all* (side+under) – *asuma* 'lie'. This is rather expected given that typically the PN of *külje all* (side+under) refers to a region and the complex item itself expresses physical proximity. The usages with the second most frequent collocate *pugema* 'creep' were also considered to belong to the unproductively formed examples because *külje alla pugema* 'to get in with somebody' can be considered to be a fixed expression. The examples formed with these two verb lemmas make up 34% of the usages of *külje all* (side+under) as a complex unit. Thus, in this case too, around two thirds of the examples are formed productively. Therefore, similar to *käe all* (hand+under) discussed above, *külje all* (side+under) is rather productive as for its use with PNs and verbs.

4.6.3. *Selja taga* (back+behind)

As was mentioned in section 4.3.3 *selja taga* (back+behind) as a complex unit occurs in four different functions – SPACE-TIME, COVERTNESS, SUPPORT, and CONCEALMENT. As these functions have developed via different paths, they are analyzed separately.

4.6.3.1. SPACE/TIME

The spatio-temporal *selja taga* (back+behind) (*n* = 4179) co-occurs with 150 PN lemmas and 112 verb lemmas.[60] As was the case for other studied phrases, the vast majority of these lemmas occur only a few times. The data show that 95% of PN lemmas and 90% of verb lemmas occur only up to 5 times. Similarly to the phrases *käe all* (hand+under) and *külje all* (side+under), the spatio-

---

[60] The number of PN lemmas is considerably smaller than in the case of the previously discussed phrases because the spatio-temporal *selja taga* (back+behind) is mostly realized as an adverb (92%), not a postposition (see section 4.4). The relatively small number of verb lemmas is related to the fact that the spatio-temporal *selja taga* (back+behind) is prone to be used without the verb. In this function, the verb is omitted on 632 occasions, which is approximately ten times more than in case of other phrases and functions of *selja taga* (back+behind).

temporal *selja taga* (back+behind) frequently occurs with proper names. In the case of spatio-temporal *selja taga* (back+behind), proper names make up just above one fourth (26%) of the PNs, which may increase the amount of low-frequency lemmas. However, after excluding the examples with proper names, the proportion of the PN lemmas that occur up to ten times remains over 90%.

4.6.3.1.1. The strongest collocates of the spatio-temporal *selja taga* (back+behind)

Table 21 shows the strongest collocates of the spatio-temporal *selja taga* (back+behind), their frequency in the etTenTen corpus, the number of times they occur with *selja taga* (back+behind) as a complex unit, and their log-likelihood score. The lemmas are ranked based on their log-likelihood score. The strongest collocates are separated with the dotted line.

Table 21. The strongest PN collocates of the spatio-temporal *selja taga* (back+behind)

PN lemma	meaning	*n* lemma (in etTenTen)	*n* collocate	log-likelihood score
meie	we	737,081	28	194
enda	own	971,567	29	187
nemad	they	178,746	20	182
tema	s/he	1,765,208	32	174
vägi	forces	19,985	8	93
kaitseliin	defense line	694	5	87
võrulane	habitant of Võru	2,158	4	59
vaenlane	enemy	14,719	5	57
liider	leader	14,726	5	57
mina	I	1,865,803	12	41
oma	own	1,346,491	10	37
meeskond	team	50,775	4	34
Saarepuu	Saarepuu*	898	2	30
nelik	quartet	903	2	30
Ansip	Ansip*	24,257	3	28
junior	junior	4,507	2	24
BMW	BMW*	5,933	2	23
eestlane	Estonian	94,866	3	20
venelane	Russian	30,708	2	16

* Proper name

The most frequent lemmas – *tema* (32 instances), *meie* (28 instances) and *enda* (29 instances) – also have the strongest statistical association with the spatio-

temporal *selja taga* (back+behind). The log-likelihood scores of these lemmas are 174, 187 and 194 respectively. In addition, *nemad* 'they', which occurred 20 times as a collocate has log-likelihood score of 182, based on which it is considered to belong to the group of the strongest collocates of spatio-temporal *selja taga* (back+behind).

As mentioned in section 4.5.1.3, the postpositional uses of the spatio-temporal function are mostly examples where *selja taga* (back+behind) carries an ordinal meaning. Thus, although the strongest collocates represented in the table occur with temporal and locative uses too, they are most representative of ordinal uses. For example, *tema* 's/he' predominantly (in 28 out of 32 instances) co-occurs with an ordinal use *selja taga* (back+behind) (see example (172)). Similarly, two of the other strong collocates – *enda* 'own' and *nemad* 'they' – are mostly (17 out of 29 and 14 out of 20 instances respectively) used with an ordinal *selja taga* 'back+behind' (see examples (173) and (174)).

(172) *H-vahegrupi-ø*   *võidu-ø*   *noppi-s*   *Tabasalu-ø*   *PK*
interim group H-GEN   win-GEN   pick-PST.3SG   Tabasalu-GEN   BC
*"Kiirus", **tema-ø** **seljataga**   Kontsertlavad.ee,   kes*
Speed   s/he-GEN   back.behind.LOC   Kontsertlavad.ee   who
võrdse-te   *võitu-de*   *korral*   *jä-i*   *Tabasalu-le*
equal-PL.GEN   victory-PL.GEN   in case   remain-PST.3SG   Tabasalu-ALL
omavahelise-s   *kohtumise-s*   *alla*   *76:82*.
rivalry-INE   meeting-INE   under   76:82
Lit. The interim group H was won by Tabasalu BC Kiirus, behind their back Kontsertlavad.ee, who, in case of equal victories was beaten by Tabasalu 76:82 in the match between them.
'The interim group H was won by Tabasalu BC Kiirus, coming in behind them Kontsertlavad.ee, who, in case of equal victories was beaten by Tabasalu 76:82 in the match between them.' [www.kossuliiga.ee]

(173) *Terav*   *vasakäär*   *on*   *korraliku-ø*   *väljakunägemise-ga*   *ja*
sharp   left edge   be.3SG   decent-GEN   appearance-COM   and
karismaatiline liidritüüpi mängija, kes jätt-is treeneri-te
charismatic   leader type   player   who   leave-PST.3SG   trainer-PL.GEN
valiku-s **enda-ø** **seljataha** *isegi koondise-ø*
choice-INE   own-GEN   back.behind.LAT   even   national team-GEN
vaatevälja-s   *ol-nud*   *Siim Lutsu-ø*.
field of view-INE   be-PST.PTCP   Siim   Luts-GEN
Lit. The sharp left edge is a charismatic leader type player with a decent appearance, who in the coaches' choice left behind his back even Siim Luts, a prospective player for the national team.
'The sharp left edge is a charismatic leader type player with a decent appearance, who was preferred by the coaches even to Siim Luts, a prospective player for the national team.' [soccernet.ee]

(174) *Viienda-l   tabelirea-l   asu-v   JK Sillamäe-ø   Kalev*
fifth-ADE   position-ADE   be-PTCP   FC Sillamäe-GEN   Kalev
reisi-b   Tartu-sse,   et   kohtu-da   sealse-ø   JK
travel-3SG   Tartu-ILL   that   meet-INF   local-GEN   FC
*Tammeka-ga,   kes   tabeli-s   vahetult   **nen-de**   selja-ø*
Tammeka-COM   who   table-INE   directly   they-PL.GEN   back-GEN
taga   *asu-b*.
behind.LOC   be-3SG
Lit. FC Sillamäe Kalev, currently on the fifth position, is travelling to Tartu to
face the local FC Tammeka who is right behind their back on the leaderboard.
'FC Sillamäe Kalev, currently on the fifth position, is travelling to Tartu to face
the local FC Tammeka who is right behind them on the leaderboard. '
[www.spordikalender.ee]

However, not all of the strong collocates are associated with the ordinal *selja
taga* (back+behind). For instance *meie* 'we', which is also strongly associated
with *selja taga* (back+behind) occurs equally frequently (ordinal 11, temporal
12, and locative 5 instances) with the temporal uses (see example (175)). When
observed individually, the locative *selja taga* (back+behind) does not have a
strong association with any of the four strongest collocates. However, the lem-
mas *vägi* 'army, force' (see example (176)), *kaitseliin* 'defense line' (see
example (177)), and *vaenlane* 'enemy' (see example (178)), which occur on 8
and 5 and 5 occasions respectively are all used with *selja taga* (back+behind)
only in the locative sense.

(175) *Meie-ø   **selja-ø**   **taha**   on   jää-nud   meeletu*
we-GEN   back-GEN   behind.LAT   be.3SG   stay-PST.PTCP   wild
XX   sajand.
XX   century
Lit. The wild 20th century has been left behind our backs.
'We have left behind us the wild 20th century.' [www.ulme.ee]

(176) *Ta   või-s   isegi   mitte   oma-ø   teene-i-d   otseselt*
s/he   can-PST.3SG   even   not   own-GEN   favor-PL-PRT   directly
pakku-da,   vaid   lihtsalt   juhata-da   lähenevad   Rootsi-ø
offer-INF   but   simply   guide-INF   approaching   Swedish-GEN
väe-d   kas   üle   veskisilla-ø   või   mujalt   läbi
force-PL   either   over   mill bridge-GEN   or   elsewhere   through
*soo-ø   **Vene-ø**   **väge-de**   **selja-ø**   **taha***.
swamp-GEN   Russian-GEN   force-PL.GEN   back-GEN   behind.LAT
Lit. S/he might not have directly offered his/her services but just guided the
approaching Swedish troops over the mill bridge or elsewhere through the
swamp to behind the back of the Russian troops.
'S/he might not have directly offered his/her services but just guided the
approaching Swedish troops over the mill bridge or elsewhere through the
swamp behind the Russian troops.' [bhr.balanss.ee]

(177) *Imelise-ø*   *väravavõimaluse-ø*   *raiska-s*   *Tani, suurepärane*
amazing-GEN   goal opportunity-GEN   waste-PST.3SG   Tani   wonderful
sööt   *keskvälja-lt*   **_kaitseliini-ø_**   **_seljataha_**,   *Tani*
pass   midfield-ABL   defense line-GEN   back.behind.LAT   Tani
kontrolli-s   *palli-ø*   *hästi*   *ette*   *ja*   *lõ-i*
controle-PST.3SG   ball-PRT   very   ahead   and   kick-PST.3SG
kiirustatult   *peale –*   *värava-st*   *mööda*.
hurryingly   on   goal-ELA   by
Lit. An amazing goal opportunity was waisted by Tani, a wonderful pass from
the midfield to behind the back of the defense line, Tani had good control of the
ball and he made a hurried go for the goal – missing the goal.
'An amazing goal opportunity was waisted by Tani, a wonderful pass from the
midfield behind the defense line, Tani had good control of the ball and he made a
hurried go for the goal – missing the goal.' [fcelva.ee]

(178) *Mõis*   *valluta-ta-kse*   *ägeda-ø*   *lahingu-ga*   *ja*   *seejärel*
manor   conquer-IMPS-PRS   bitter-GEN   battle-COM   and   thereafter
tungi-ta-kse   *üle*   *Rauna-ø*   *jõe-ø*   *kaugele*   **_vaenlase-ø_**
invade-IMPS-PRS   over   Rauna-GEN   river-GEN   far   enemy-GEN
selja-ø   **_taha_**.
back-GEN   behind.LAT
Lit. The manor will be conquested in a bitter battle and thereafter they will
invade far behind the back of the enemy, over the river Rauna.
'The manor will be conquested in a bitter battle and thereafter they will invade
far behind the enemy, over the river Rauna.' [www.hot.ee]

Based on these examples, it seems that the locative use of *selja taga*
(back+behind) prefers collective PNs. This is connected to the fact that the col-
lective PN is perhaps the clearest parameter to distinguish the complex locative
postposition *selja taga* (back+behind) from the freely combined simple post-
positional phrase (see section 4.3.3). The other frequent PNs do not form any
particular semantic class. This may be due to the fact that the most frequent
collocates are pronouns, which have a very general reference.

Similar to *käe all* (hand+under) and *külje all* (side+under) that were dis-
cussed previously, the combination of *selja taga* (back+behind) and its most
frequent PNs do not form holistic expressions that would carry distinct phrasal
meanings. However, such utterances are considered more or less fixed units
and, therefore, as not contributing to the productivity of the spatio-temporal use
of *selja taga* (back+behind).

In the following, I present and analyze the verbs that have the strongest sta-
tistical association with the spatio-temporal *selja taga* (back+behind).

Table 22. The strongest verbal collocates of the spatio-temporal *selja taga* (back+behind)

| verb lemma	meaning	*n* lemma in etTenTen	*n* collocate	log-likelihood score
jääma/jätma	remain/leave	648,139	1,588	77,955
olema	be	9,559,606	1,686	28,857
hoidma/hoiduma	keep/keep out of	151,324	43	960
vaatama	look	265,953	29	354
pilku heitma	cast a glance	563	4	255
platseeruma	keep a position	609	3	176
juhatama	lead	14,329	5	123
asuma	lie	118,320	9	79
tungima	force	9,190	4	109
lõppema/lõpetama	end/finish	145,431	8	48
startima	start	5,281	3	91
heitma	cast	21,152	4	71
pääsema	get through	47,205	5	60
jagunema	distribute	8,362	3	75
pilke saatma	glance	41	1	82
püsima	stay	32,817	3	32
jõudma	reach	233,096	7	12
heljuma	float	267	1	55
kemplema	dispute	275	1	54
kaduma	disappear	75,790	4	23

The strongest collocates of the spatio-temporal *selja taga* (back+behind) are *jääma/jätma* 'remain/leave' and *olema* 'be'. *Jääma/jätma* 'remain/leave' is occurs 1588 times. With the log-likelihood score of 77,955 it is without a doubt the strongest collocate of the spatio-temporal *selja taga* (back+behind). The high frequency (and the strong association) is also connected with the fact that *jääma/jätma* actually includes two lemmas. However, the usages with *jääma* 'remain' (as in (179)) and *jätma* 'leave' (as in (180)) are very close, so they are taken to represent the same usage (see principle iii in section 4.6). The verb *olema* 'be' is even more frequent than *jääma/jätma* 'remain/leave' but due to the fact that it is a highly frequent verb in general, it receives a much lower association score (28,857). However, relative to other lemmas, *olema* 'be' is still much more strongly associated with *selja taga* (back+behind), see example (181). The rest of the verbs are much less frequent (43 and 29 occurrences) and their association scores are many times lower than that of olema 'be' and jääma/jätma 'remain/leave'.

(179) ***Selja-ø*** ***taha*** ***jä-i*** *viis* *aasta-t* *töö-d*
back-GEN behind.LAT remain-PST.3SG five year-PRT work-PRT
California-s.
California-INE
Lit. Five years of work in California was left behind the back.
'Five years of work in California was left behind.' [www.sirp.ee]

(180) *Taha-ks* *loot-a,* *et* *Kaukaasia-ø* *ajaloo-s* *on*
want-COND hope-INF that Caucasia-GEN history-INE be.3SG
traagiline *lehekülg* *lõplikult* ***selja-ø*** ***taha***
tragic page for good back-GEN behind.LAT
jäe-tud.
leave-PST.PTCP
Lit. I would like to hope that the tragic page in the history of Caucasia has been
left behind the back for good.
'I would like to hope that the tragic page in the history of Caucasia has been left
behind for good.' [arvamus.postimees.ee]

(181) *Euroopa-ø* ***selja-ø*** ***taga*** ***on*** *aastatuhande-d*
Europe-GEN back-GEN behind.LOC be.3SG millennium-PL
kultuuri-ø *ajalugu-ø,* *vaimsuse-ø* *kroonika-t*.
culture-GEN history-PRT spirituality-GEN chronicle-PRT
Lit. Europe has millenniums of history of culture, chronicles of spirituality
behind its back.
'Europe has millenniums of history of culture, the chronicles of spirituality under
its belt.' [www.kirikiri.ee]

### 4.6.3.1.3. The productive use of the spatio-temporal *selja taga* (back+behind)

Neither the strongly associated PN lemmas or the verb lemmas are considered to be formed productively. In the following, the amount of the productively formed examples are observed. Figure 18 shows the cumulative percentage of the instances of the spatio-temporal uses of *selja taga* (back+behind) based on the frequency of the PN lemmas.
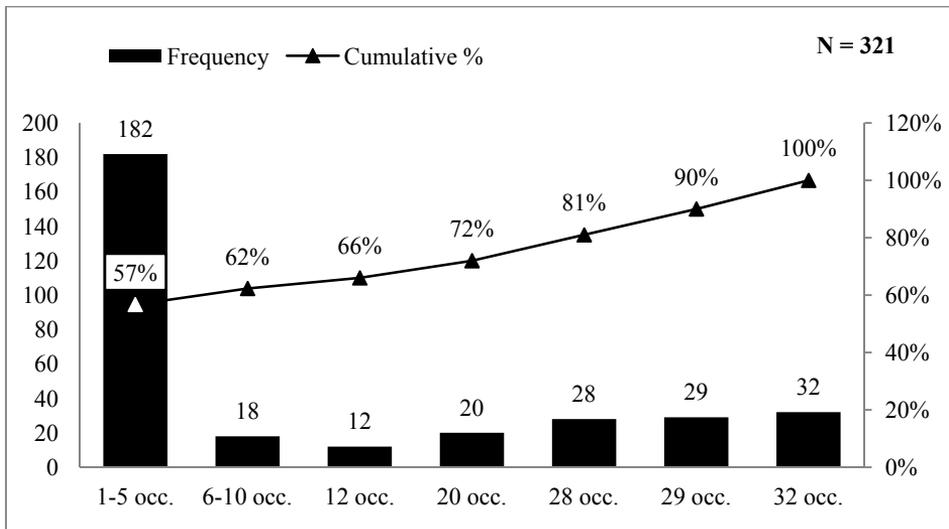
Figure 18. The cumulative percentage of examples of spatio-temporal *selja taga* (back+behind) formed with PN lemmas based on the number of occurrence

The data show that about two-thirds (66%) of the examples with spatio-temporal use of *selja taga* (back+behind) are formed with low-frequency lemmas. It can be observed in Figure 18 that 57% of the examples (182 out of 321) are formed with PN lemmas that appeared only 1–5 times in the sample. The lemmas that occur up to ten times accounted for 62% (200 out of 321) of the data and those that occur up to 12 times 66% of the data. The rest of the examples (34%) are formed with lemmas that occur on at least 20 occasions and that were, based on the log-likelihood score (see Table 21), considered to be strong collocates of the spatio-temporal use of *selja taga* (back+behind). Thus, it is concluded that based on the use of the PN lemmas, the spatio-temporal *selja taga* (back+behind) is quite productive. However, excluding the examples with proper names reduces the proportion of examples with low-frequency lemmas to 54%. Thus, it is concluded that the productivity is somewhat dependent on the fact that the spatio-temporal *selja taga* tends to co-occur with proper names.

Figure 19 shows the cumulative percentage of the instances of the spatio-temporal *selja taga* (back+behind) as complex unit based on the frequency of the verb lemma.
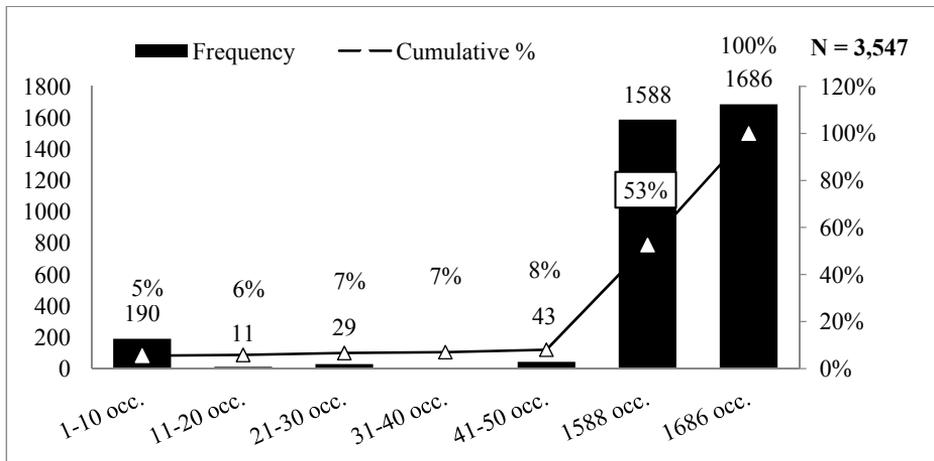
Figure 19. The cumulative percentage of examples of the spatio-temporal *selja taga* (back+behind) formed with verb lemmas based on the number of occurrence

The data show that the use of spatio-temporal *selja taga* (back+behind) is rather restricted as for the verb lemmas with which it co-occurs. The two strongest collocates *jääma/jätma* 'remain/leave' and *olema* 'be' which occurred on 1588 and 1686 occasions respectively. Examples with these verbs account for most (92%) of the usages of the spatio-temoral function. Thus, the proportion of examples that are formed with low-frequency lemmas and, therefore, productively is only 8%. Thus, it seems that the spatio-temporal interpretation of *selja taga* (back+behind) is still depenedent on the meaning of the verb. This is in line with Lehmann's (1991: 503) observations on 'new-vawe' German complex prepositions, which arised in 1970s – *im Zuge* 'by, during, in', *im Wege* 'by (way/means of)', *im Vorfeld* 'on the eve (of), before', etc. His data showed that the newly formed complex prepositions are at first used in rather restricted contexts, i.e. with certain collocations or even phraseologisms. According to Lehmann, such new usages cannot be understood without the supporting context in which they were originally coined (Lehmann 1991: 503). As will be demonstrated in section 4.8.4.2, the history of spatio-temporal *selja taga* (back+nehind) goes back to much earlier times than 1970s. Nevertheless, the verbs *olema* 'be' and *jääma/jätma* 'remain/leave' are still central in using and unserstanding the spatio-temporal meaning of *selja taga* (back+behind).

4.6.3.1.3. Summary of the productivity of the spatio-temporal *selja taga* (back+behind)

The spatio-temporal *selja taga* (back+behind) co-occurs with 150 PN lemmas and 112 verb lemmas. Most of these are low-frequency lemmas – 95% of the PN and 90% verb lemmas occur up to five times.

Based on absolute occurrences and log-likelihood scores, the spatio-temporal *selja taga* (back+behind) has four strong collocates – *meie* 'we', *enda* 'own',

tema 's/he' and *nemad* 'they'. Usages formed with these PN lemmas are con-sidered to stand for more or less fixed expressions. Such usages make up about one third of all the examples. Thus, because about two thirds of the examples are formed with weak collocates, the spatio-temporal *selja taga* (back+behind) could be considered rather productive. However, when examples with proper names are excluded from the data, the amount of productive uses dropped to 54%. Morover, the analysis of the verb lemmas suggests that the use of the spatio-temporal *selja taga* (back+behind) is quite restricted. There are two extremely frequent collocates *jääma/jätma* 'remain/leave' and *olema* 'be', which are used in 92% of the examples. This does not suggest productivity as a complex unit. Furthermore, the fact that the spatio-temporal function is strongly inclined to be used as an adverb indicates that the development of complex postpositional spatio-temporal *selja taga* (back+behind) is considered to still be in its early stages.

4.6.3.2. COVERTNESS

As was shown in section 4.2, *selja taga* (back+behind) is used to expresses COVERTNESS in 870 instances. Within these 870 examples, *selja taga* (back+behind) takes 129 PN lemmas and 294 verb lemmas. 93% of the PN lemmas and 91% of the verb lemmas occur up to 5 times. As in case of other phrases (or individual functions) discussed above, the more frequent lemmas make up only a small percentage (4%) of the lemmas. These will be observed more closely in the next section.

4.6.3.2.1. The strongest collocates of *selja taga* (back+behind) expressing COVERTNESS

The strongest PN lemmas of *selja taga* (back+behind) along with their frequency in the etTenTen corpus, the number of times they occur with *selja taga* as a complex postposition, and their log-likelihood score are presented in Table 23.

Table 23. The strongest PN collocates of *selja taga* (back+behind) in the function COVERTNESS

PN lemma	meaning	*n* lemma (in etTenTen)	*n* collocate	log-likelihood score
tema	s/he	1,765,208	56	343
nemad	they	178,746	34	330
rahvas	people	141,245	29	286
mina	I	1,865,803	48	274
sina	you	645,915	24	155
omanik	owner	58,687	10	95
meie	we	737,081	17	93

PN lemma	meaning	*n* lemma (in etTenTen)	*n* collocate	log-likelihood score
teie	you.PL	182,123	10	72
avalikkus	public	21,101	6	63
minister	minister	30,534	5	47
tööandja	employer	44,860	5	43
õpetaja	teacher	125,700	4	25
keegi	someone	267,117	4	19
ema	mother	107,212	3	18
mees	man	303,390	4	17
riik	state	356,114	4	16
juhendaja	supervisor	9,499	1	9
juht	leader	257,232	2	7
teine	other	717,260	3	6

When expressing COVERTNESS, *selja taga* (back+behind) does not form any distinct idiomatic expressions with its stronger collocates – *tema* 's/he', *nemad* 'they', *rahvas* 'people', *mina* 'I', and *sina* 'you' – , rather, the utterances formed with these lemmas exemplified in (182)–(186) are considered to be more prototypical examples of *selja taga* (back+behind) in this function. Based on the most frequent collocates in Table 23, there seem to be no particular semantic restrictions to the PN lemmas which *selja taga* (back+behind) co-occurs with in this function.

(182) *Ta   ei   salli-ø,   kui   keegi   **tema-ø**   **selja-ø**
        s/he   NEG   tolerate-CONNEG   if   someones   s/he-GEN   back-GEN
        taga   tema-ø   teh-tud   töö-d   kontrolli-b.
        behind.LOC   s/he-GEN   do-PST.PTCP   work-PRT   controle-3SG
        'S/he cannot stand it if someone checks his/her work behind his/her back.'
        [www.rajaleidja.ee]

(183) *Linnavoliniku-d   on   hämmeldunud,   et   **nen-de**
        city councilman-PL   be.3PL   baffled   that   they-PL.GEN
        seljataga   allkirjasta-s   Ivi   Eenmaa   Bulgaaria-ø
        back.behind.LOC   sign-PST.3SG   Ivi   Eenmaa   Bulgaria-GEN
        linna-ga   koostöökokkuleppe-ø*.
        c city-COM   ooperation agreement-GEN
        'The city councilmen are baffled that Ivi Eenmaa signed a cooperation
        agreement with a Bulgarian city behind their backs.' [www.lounaleht.ee]

(184) *Selge,   et   mäng käi-b   jälle   **rahva-ø   selja-ø   taga**,*
obvious that game go-3SG again people-GEN back-GEN behind.LOC
sest   eesti   rahvas   on   ju   loll   ja   anna-b
because Estonian nation   be.3SG after all stupid and   give-3SG
kõik   ande-ks.
everything   excuse-TRL
'It is clear that it is going on behind the back of the people because after all,
Estonians are stupid and forgive everything.' [www.epl.ee]

(185) *Usalda-si-n   oma-ø   saladuse-ø   klassijuhataja-le,   kes*
trust-PST-1SG   own-GEN   secret-GEN   homeroom teacher-ALL   who
ol-I   ka   psühholoog,   aga   tema   kuuluta-s
be-PST.3SG   also psychologist but   s/he   blat out-PST.3SG
minu-ø   selja-ø   taga   *se-da   kogu   klassi-le*
I-GEN   back-GEN   behind.LOC   this-PRT   whole class-ALL
ning kutsu-s   poisi-ø   vanema-d   kooli-ø …
and   call-PST.3SG   boy-GEN   parent-PL   school-ILL
'I trusted my secret to my class teacher who also happened to be the psychologist
but s/he blatted it out to the whole class behind my back and called the boy´s
parents to school.' [www.caritas.ee]

 (186) *Ka   kõige   pare-m   sõber, kes   paista-b   ole-vat   su-ø*
also   most   good-COMP friend who look-3SG   be-QUOT   you-GEN
*kaasvõitleja, või-b   **su-ø   selja-ø   taga**   kogu*
cofighter   might-3SG   you-GEN back-GEN   behind.LOC   whole
ettevõtte-ø   maha müü-a.
company-GEN   sell out-INF
'Even your best friend who seems to be your cofighter might sell out the whole
company behind your back.' [www.director.ee]

The strongest verbal collocates of *selja taga* (back+behind) in the function
COVERTNESS are *rääkima* 'talk' and *tegema* 'do' (see Table 24). *Rääkima*
'talk' co-occurs with *selja taga* (back+behind) 87 times and has the log-likeli-
hood score of 774, and is its strongest collocate. A close second is the lemma
tegema 'do' which occurs on 93 occasions and whose association score is 614.
The rest of the lemmas are considered to have a weaker association with *selja
taga* (back+behind) because the log-likelihood values are considerably lower
(326 and less) as are the absolute frequencies of these lemmas.

Table 24. The strongest verbal collocates of *selja taga* (back+behind) in the function COVERTNESS

verb lemma	meaning	*n* lemma in etTenTen	*n* collocate	log-likelihood score
rääkima	talk	324,712	87	774
tegema	do	1,091,109	93	614
kiruma	curse	4,253	22	326
sosistama	whisper	2,559	17	260
toimuma	happen	252,280	28	200
naerma	laugh	21,385	16	175
otsustama	decide	109,727	18	142
klatšima	gossip	102	7	140
suskima	cabal	107	7	139
ründama	attack	12,843	11	123
itsitama	giggle	1,268	8	122
kallale tungima	assault	934	6	92
ajama	carry (out)	85,380	12	91
tegutsema	take action	55,912	11	91
mõnitama	taunt	3,436	7	91
halvustama	derogate	2,220	6	81
irvitama	make fun (of)	2,635	6	79
sahkerdama	swidle	346	4	66
susisema/susistama	hiss	349	4	66
vastu võtma	accept	64,571	8	59

Thus, based on the verb lemma, *selja taga rääkima* 'talk behind one's back' (see example (187)) and *selja taga [midagi] tegema* 'do [something] behind one's back' (see example (188)) are the most typical usages of *selja taga* (back+behind) in this function. However, such usages are not idiomatic expressions; *selja taga* (back+behind) carries quite a similar meaning in examples that are formed with weaker lemmas. Nevertheless, it should be noted that *tegema* 'do' often acts as a light verb, i.e. it is combined with another word, such as *halba tegema* 'do harm' or *lepinguid tegema* 'make contarcts'. It may also act as a pro-verb, i.e. referring to actions that have been mentioned before: (*tihti tehakse seda sinu selja taga* 'often it is done behind your back').

(187) *Mi-da   na-d   mõtle-vad   kuidas   tegutse-vad,   mi-da*
     what-PRT   they-PL   think-3PL   how   take action-3PL   what-PR
     *su-ø   **selja-ø   taga   räägi-vad**,   draama…   mitte*
     you-GEN   back-GEN   behind.LOC   talk-3PL   drama   not
     miski   ei   puuduta-ø   sin-d!
     nothing   NEG   touch-CONNEG   you-PRT
     'What do they think, how do they act, what do they say behind your back,
     drama… nothing bears on you.' [www.hot.ee]

(188) *Te-i-l   ei   ole-ø   aimu-gi   mi-da   kõike-ø*
     you-PL-ADE   NEG   be-CONNEG   idea-CL   what-PRT   all-PRT
     teie-ø   ontlik   ja   hästi   kasvata-tud   laps   teie-ø
     you-GEN   neat   and   well   behave-PST.PTCP   kid   you-GEN
     ***selja-ø   taga   teh-a**   või-b.*
     back-GEN   behind.LOC   do-INF   might-3SG
     'You have no idea what your neat and well raised child can do behind your
     back.' [rahvahaal.delfi.ee]

Although the strongest (pro)nominal collocates of *selja taga* (back+behind) did not represent a single semantic group in this function, there are similarities between the strongest verb lemmas. It can be observed in Table 24 that the lemmas often refer to various speech acts – e.g. *sosistama* 'whisper', *kiruma* 'curse', klatšima 'gossip'(see example (189)). Another noticeable trait of the verbs is that they carry a negative attitude towards the LM, for instance *halvustama* 'derogate', *suskima* 'cabal', *mõnitama* 'taunt' (as in (190)). This is, of course, connected to the function that *selja taga* (back+behind) carries.

(189) *Aga   kui   viisakas   on   **seljataga**   pärast   **klatsi-da**,   et*
     but   how   polite   be.3SG   back.behind.LOC   after   gossip-INF   that
     vaat   kus   põrsas,   tõ-i   mu-lle   oma-ø   vana-ø   kraami-ø!
     look   what   piglet   bring-PST.3SG   I-ALL   own-GEN   old-PRT   stuff-PRT
     'But how polite is it to gossip behind one's back afterwards and say what a pig,
     s/he brought me his/her old stuff!' [rahvahaal.delfi.ee]

(190) *Kujuta ette,   Lust ,   kuidas   IT-Ø   poisi-d   sin-d   **selja-ø***
     imagine   Lust   how   IT-GEN   boy-PL   you-PRT   back-GEN
     ***taga   mõnita-vad**.*
     behind.LOC   taunt-3PL
     'Imagine, Lust, how the IT-guys are taunting you behind your back.'
   [naistekas.delfi.ee]

4.6.3.2.2. The productive use of *selja taga* (back+behind) expressing COVERTNESS

Although these strongest collocates do not form distinct idiomatical phrases with *selja taga* (back+behind), these examples that are formed with the strong-

est collocates are not considered to stand for the productive use of the complex unit. Thus, it is useful to determine the amount of examplesproductively formed examples.

Figure 20 shows the cumulative percentage of the instances where *selja taga* (back+behind) is used as a complex postposition that expresses COVERTNESS based on the frequency of the PN lemma.
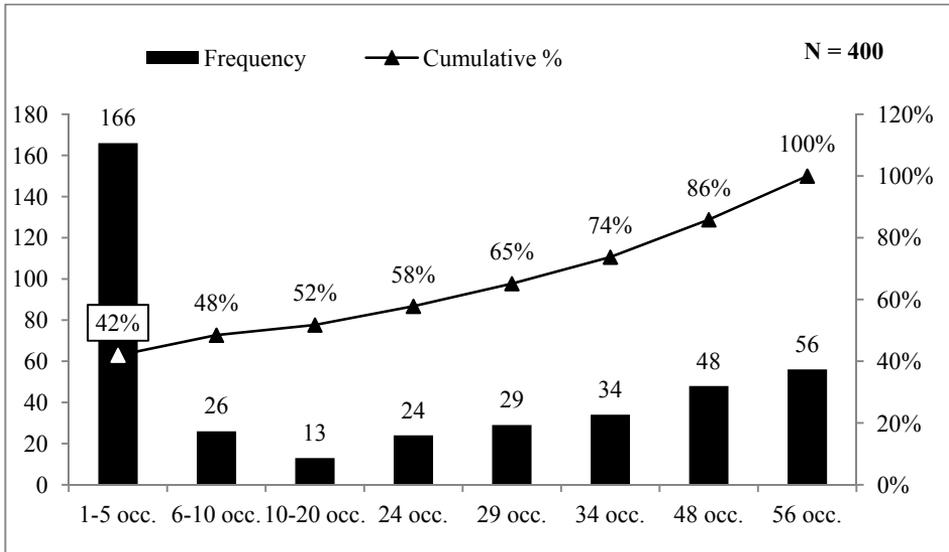
Figure 20. The cumulative percentage of examples of *selja taga* (back+behind) in the function COVERTNESS formed with PN lemmas based on the number of occurrence

The data show that just over half (52%) of the examples are formed pro-ductively. Figure 20 shows that 41% of the examples (166 out of 400) are formed with PN lemmas that appear 1–5 times in the data, 48% (192 examples out of 400) of the PN lemmas occur up to ten times and 52% occur up to 20 times (205 examples out of 400). The rest of the examples (48%) are formed with the lemmas that were shown to be strong collocates of *selja taga* (back+behind) in this function (see Table 23). Thus, in this aspect, the use of *selja taga* (back+behind) as a complex postposition that expresses COVERT-NESS is not as productive as the other functions of complex postpositional *selja taga* (back+behind) discussed in the previous section and in the following sec-tions (4.6.3.1. and 4.6.3.4).

Figure 21 shows the cumulative percentage of the instances of COVERTNESS based on the frequency of the verb lemma. The data suggests that examples that include the strongest collocates *rääkima* 'talk' and *tegema* 'do', make up 21% of the data, i.e. 79% of the examples of *selja taga* (back+behind) expressing COVETRNESS are formed productively.
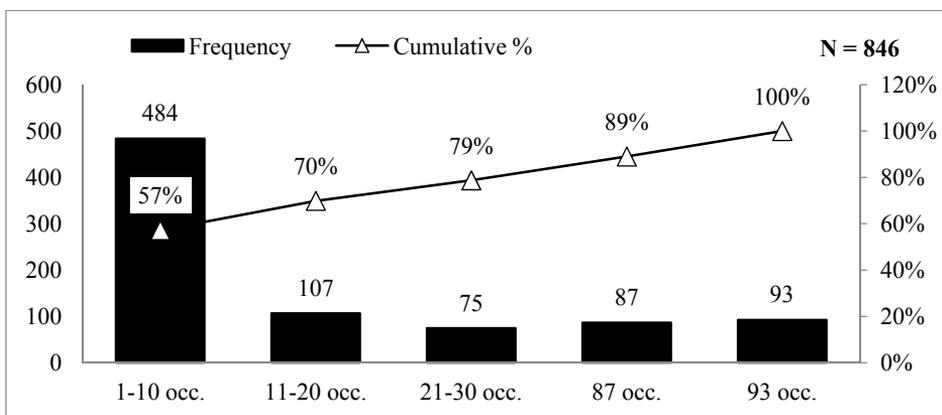
Figure 21. The cumulative percentage of examples of *selja taga* (back+behind) expressing COVERTNESS formed with verb lemmas based on the number of occurrence

4.6.3.2.3. Summary of the productivity of *selja taga* (back+behind) expressing COVERTNESS

When expressing COVERTNESS *selja taga* (back+behind) co-occurs with 129 PN lemmas and 294 verb lemmas. Although most of the lemmas are low-frequency collocates of *selja taga* (back+behind) – 93% the PN lemmas and 91% of the verb lemmas occur only up to 5 times – there are some frequent collocates, which are suspect of non-productive usage.

The most frequent PN lemmas (*tema* 's/he', *nemad* 'they', *rahvas* 'people', *mina* 'I', and *sina* 'you') make up just under half of the examples. This suggests that when expressing COVERTNESS, the use of *selja taga* (back+behind) is rather restricted. However, regarding its use with the verbs, the examples with strong collocates (*rääkima* 'talk' and *tegema* 'be') make up 21% of all the examples of COVERTNESS. The rest of the examples (79%) are formed productively.

4.6.3.3. SUPPORT

The total number of examples where *selja taga* (back+behind) is used to express SUPPORT is 610. Within these examples, *selja taga* (back+behind) takes 202 PN and only 62 verb lemmas. Quite expectedly, most of the lemmas are rather infrequent – 95% of the PN lemmas and 84% of lemmas occur up to 5 times.[61] As in the cases discussed previously, the very frequent lemmas make up a very small proportion of all the PN lemmas. These are observed in greater detail in the next section.

---

[61]   When expressing the function SUPPORT, *selja taga* (back+behind) also frequently occurs with proper names. Proper names make up just below one fourth (23%) of the examples. However, excluding these examples does not affect the amount of low-frequency lemmas.

4.6.3.3.1. The strongest collocates of *selja taga* (back+behind) expressing SUPPORT

Table 25 shows that the strongest PN collocates of *selja taga* (back+behind) in this function are *nemad* 'they', *tema* 's/he', and *enda* 'own'. The log-likelihood scores of these lemmas are 456, 352, and 178 respectively. The rest of the lemmas – starting from *oma* 'own' which occurred on 24 occasions – have lower log-likelihood values (≤110), and are considered to have a weaker association with *selja taga* (back+behind).

Table 25. The strongest PN collocates of *selja taga* (back+behind) in the function SUPPORT

PN lemma	meaning	*n* lemma (in etTenTen)	*n* collocate	log-likelihood score
nemad	they	178,746	46	456
tema	s/he	1,765,208	60	352
enda	own	971,567	31	178
oma	own	1,346,491	24	110
sina	you	645,915	18	98
kes	who	1,036,055	19	88
mina	I	1,865,803	20	71
Obama	Obama*	7,028	5	60
teie	you.PL	182,123	9	60
meie	we	737,081	12	53
demokraat	democrat	2,953	3	38
Savisaar	Savisaar*	21,627	4	37
Serbia	Serbia*	3,674	3	37
kommunist	communist	7,149	3	33
usa	USA*	72,983	4	27
ansip	Ansip*	24,257	3	25
poliitik	politician	35,535	3	23
arst	doctor	82,849	3	18
valistsus	government	107,308	3	16
laps	child	475,173	4	12

* Proper name

When expressing SUPPORT, *selja taga* (back+behind) does not form any distinct idiomatic expressions with its stronger collocates, rather, the utterances *tema selja taga* (s/he-GEN back+behind) (see example (191)), *nende selja taga* (they-GEN back+behind) (see example (192)), and *enda selja taga* (own-GEN back+behind) (see example (193)) are considered to be more prototypical examples of complex postpositional *selja taga* (back+behind) in this function.

(191) *Lisaks   selle-le   ol-i   ta   haka-nud   ka*
besides   this-ALL   be-PST.3SG s/he   begin-PST.PTCP also
*luterlase-ks   ja   nõnda kukku-s   paavsti-ø   toetus   **tema-ø***
Lutheran-TRL   and   so   fall-PST.3SG pope-GEN support   s/he-GEN
selja-ø   tagant   kohe   ära.
back-GEN   behind.SEP   straight away   off.
Lit. In addition to that, s/he had also become a Lutheran and thus the support of
the pope fell off from behind his/her back straight away.
'In addition to that, s/he had also become a Lutheran and thus the pope was no
longer behind his/her back.' [et.wikipedia.org]

(192) *Endine   Arkansase-ø   kuberner   hinda-s   ka,   et   demokraat*
former   Arkansas-GEN   governor   assess-PST.3SG   also that democrat
Barack Obama   ei   vali-ø   enda-ø   valimiskampaania-ø
Barack Obama   NEG   choose-CONNEG   own-GEN campaign-GEN
*partneri-ks   Hillary Clintoni-t,   kuna   **nen-de   selja-ø***
partner-TRL   Hillary Clinton-PRT   because they-PL.GEN   back-GEN
taha   koondu-nud   inimese-d   ei   saa-ø
behind.LAT   consolidate- PST.PTCP   people-PL   NEG   get-CONNEG
eriti omavahel   läbi   ning   selline   liit   ei   pälvi-ks
much   in between along   and   this kind union   NEG   earn-COND
nen-de   heakskiitu-ø . [riigikogu.postimees.ee]
they-PL.GEN   approve-PRT
Lit. The former governor of Arkansas also thought that the democrat Barack
Obama will not be choosing Hillary Clinton as his campaign partner because the
people gathered behind the back of either of them do not get along and this union
would not have earned their approval.
'The former governor of Arkansas also thought that the democrat Barack Obama
will not be choosing Hillary Clinton as his campaign partner because the people
behind either of them do not get along and this union would not have earned
their approval.' [riigikogu.postimees.ee]

(193) *Nüüd   on   Bella-l   ja   Edwardi-l   vaja   lühikese-ø*
now   be-3SG   Bella-ADE   and   Edward-ADE   need   short-GEN
*aja-ga   koguda   **enda-ø   selja-ø   taha***   nii
time-COM   collect-INF   own-GEN   back-GEN   behind.LAT   so
liitlas-i,   kui   vähegi   võimalik,   et   kaits-ta   oma-ø
ally-PL.PRT   if   at all   possible   that   protect-INF   own-GEN
perekonda-ø   viimase-s,   kõikeotsustava-s   lahingu-s.
family-PRT   last-INE   determinant-INE   battle-INE
Lit. Now Bella and Edward need to gather as many allies as possible behind their
back with a short time to protect their family in the final, definitive battle.
'Now Bella and Edward need to gather as many allies behind them as possible
with a short time to protect their family in the final, definitive battle.'
[www.forumcinemas.ee]

It can be observed in Table 25 that *selja taga* (back+behind) is most often used to express SUPPORT in a political sense. For instance, *selja taga* is often complemented by lemmas that refer to politicians, such as *Obama* (as in example (194)), *Savisaar*, and *demokraat* 'democrat'. This is also suggested by the examples with frequently occurring PNs *tema* 's/he' (see example (191)) and *nemad* 'they' (see example (192)). In both of these examples, the pronoun actually refers to political figures. This can also be observed among the examples where the PN refers to a state such as Serbia or the USA. In such cases, the postposition is used to express the (political) support of countries (as in (195)). Nevertheless, *selja taga* (back+behind) is not confined to political discourse. It can be also used to express support in other domains. For instance, in example (196) it is used to express the (supportive) relationship between the doctor and the medical system.

(194) *USA   meediakanali-te   prognoosi-de   järgi   lähe-b*
        USA   media channel-PL.GEN   forecast-PL.GEN   according to   go-3SG
        *Pennsylvania-ø   osariik   **Obama-ø   selja-ø   taha**,   mis*
        Pennsylvania-GE   state   Obama-GEN   back-GEN   behind.LAT   what
        ole-ks   valus   löök   McCaini-ø   kampaania-le.
        be-COND   painful   kick   McCain-GEN   campaign-ALL
        Lit. According to the forecasts of the media channels in the USA, the state of
        Pennsylvania will go behind Obama's back, which would be a hard blow to
        McCain's campaign.
        'According to the forecasts of the media channels in the USA, the state of
        Pennsylvania will get behind Obama, which would be a hard blow to McCain´s
        campaign.' [usa.postimees.ee]

(195) *USA   on   juba   pikalt   toeta-nud   Kosovo-ø*
        USA   be.3SG   already   long   support-PST.PTCP   Kosovo-GEN
        albaanlas-te   Iseseisvuspüüdlus-i ,   samas   kui   Venemaa
        Albanian-PL.GEN   independence endeavor-PL.PRT   while   if   Russia
        *on   asu-nud   **Serbia-ø   selja-ø   taha.***
        be.3SG   set-PST.PTCP   Serbia-GEN   back-GEN   behind.LAT
        Lit. The USA has been supporting the pursuit of independence of the Albanians
        in Kosovo for a long time already, while Russia has taken a place behind the
        back of Serbia.
        'The USA has been supporting the pursuit of independence of the Albanians in
        Kosovo for a long time already, while Russia has began to get behind Serbia.'
        [www.vm.ee]

(196) *See   eelda-b   aga,   et   **perearsti-ø   selja-ø   taga***
        this   assume-3SG   but   that   G.P.-GEN   back-GEN   behind.LOC
        on   tugev   tugisüsteem.
        be.3SG   strong   support system
        Lit. That, however, implies that there is a strong support system behind the back
        of the G.Ps.
        'That, however, implies that there is a strong support system behind the G.Ps.'
        [www.med24.ee]

It can be observed in Table 26 that when expressing SUPPORT, *selja taga* (back+behind) often collocates with *seisma* 'stand' (119 occurrences; log-likelihood score of 1566). A close second is the lemma *olema* 'be' which occurs on 259 occasions and whose log-likelihood score is 1306. The rest of the lemmas are considered to have a weaker association with *selja taga* (back+behind) because the log-likelihood values are considerably lower (563 and less) as are the absolute frequencies of these lemmas.

Table 26. The strongest verbal collocates of *selja taga* (back+behid) in the function SUPPORT

verb lemma	meaning	*n* lemma in etTenTen	*n* collocate	log-likelihood score
seisma	stand	75,814	119	1,566
olema	be	9,559,606	259	1,306
koondama/ koonduma	consolidate / be consolidated	19,724	41	563
asuma	lie	118,320	24	218
koguma/kogunema	gather / be gathered	25,843	18	208
joonduma	align	442	3	48
tundma	feel	224,072	8	45
saama	get	1,871,918	16	44
astuma	step	64,581	5	36
toetama	support	86,145	5	33
jääma/jätma	remain/leave	648,139	8	28
minema	go	605,794	7	23
pugema	creep	6,241	2	20
rivvi võtma	line up	26	1	20
konspireerima	conspire	29	1	19
palkama	hire	7,388	2	19
istuma	sit	68,109	3	18
tunnetama	perceive	10,381	2	18
leidma	find	424,579	5	17
tooma	bring	229,994	4	17

Thus, based on the verb lemma, *selja taga seisma* 'stand behind one's back' (197) and *selja taga olema* 'be behind one's back' (198) are the most typical usages of *selja taga* (back+behind) in this function. In this case, too, the usages with the strongest collocates are semantically rather similar to that of weaker collocates, that is, they do not form idiomatic expressions. It was suggested above that this function is often used in (but not confined to) to political discourse. This is also reflected in its use with verbs – there is no one particular semantic class that is associated with this meaning but it is observable that the

most frequent verbs are associated with examples that express political support – *koondama/koonduma* 'consolidate/be consolidated' (see example (199)), *koguma/kogunema* 'gather/be gathered' (200), and *toetama* 'support' (201)).

(197) *Meie   kristlase-d   saa-me   **seis-ta**   Iisraeli-ø   eest   ja*
we   christian-PL   can-1PL   stand-INF   Israel-GEN   for   and
*Iisraeli-ø   **selja-ø**   **taga***.
Israel-GEN   back-GEN   behind.LOC
Lit. Us, Christians, can stand for Israel and behind the back of Israel.
'Us, Christians, can stand for Israel and support them.' [www.usk.ee]

(198) *Nen-de   inimes-te   **selja-ø**   **taga**   ol-i   eesti*
they-PL   people-PL.GEN   back-GEN   behind.LOC   be-PST.3SG   Estonian
rahvas.
people
Lit. The Esotnian people were behind the back of these people.
'The Estonian people were behind these people.' [syndikaat.ee]

(199) *Olümpiahooaja-l   on   Kristina   Šmiguni-ø   eesmärgi-ks*
olympic season-ADE   be.3SG   Kristina   Šmigun-GEN   goal-TRL
koonda-da**   oma-ø   **selja-ø**   **taha
consolidate-INF   own-GEN   back-GEN   behind.LAT
par-ima-te-st   par-im-ad.
good-SPL-PL-ELA   good-SPL-PL
Lit. During the Olympic season it is the goal of Kristina Šmigun to gather the best of the best behind her back.
'During the Olympic season it is the goal of Kristina Šmigun to have the best of the best behind her.' [eok.ee]

(200) *vaenlase-ø   kuju   on   loo-dud,   kõige-ø   ohtliku-m*
enemy-GEN   figure   be.3SG   create-PST.PTCP   all-GEN   dangerous-COMP
ja   suure-m   konkurent   kõrvalda-tud   ja   valija
and   big-COMP   competitor   remove-PST.PTCP   and   elector
kogune-b**   IRLi-ø   ja   reformi-ø   **selja-ø
be gathered-3SG   IRL-GEN   and   reform-party-GEN   back-GEN
taha.
behind.LAT
Lit. The figure of the enemy has been created, the most dangerous and biggest competitor has been removed and the voter is gathering behind the back of IRL and the Reform Party.
'The figure of the enemy has been created, the most dangerous and biggest competitor has been removed and the voter is supporting IRL and the Reform Party.' [www.ekspress.ee]

(201) *Kohus   arvesta-b   kindlasti   varalise-ø   kahjutekitaja-ø*
court   count-3SG   definetly   pecuniary-GEN   tortfeasor-GEN
eelneva-i-d   tegemis-i   ja   jälgi-des   senise-i-d
previous-PL-PRT   doing-PL.PRT   and   monitor-GER   current-PL-PRT
skandaalse-i-d   sündmus-i,   siis   ka   se-da,   milline   erakond
scandalous-PL-PRT   event-PL.PRT   then   also   this-PRT   which   party
*on   **selja-ø**   **taga**   **toeta-ma-s***
be.3SG   back-GEN   behind.LOC   support-SUP-INE
Lit. The court will definitely take into account the previous doings of the person
who has caused material damage and following the current scandalous events,
also which party is supporting [him/her] behind the back.
'The court will definitely take into account the previous doings of the person
who has caused material damage and following the current scandalous events,
also which party is behind [him/her].' [www.epl.ee]

Although the strongest collocates do not form distinct idiomatical phrases with
selja taga (back+behind), the examples with the strongest collocates are not
considered to amount to productive use of *selja taga* (back+behind). Thus, the
following section the amount of productively formed examples is determined.

4.6.3.3.2. The productive use of *selja taga* (back+behind) when expressing SUPPORT

Figure 22 shows the cumulative percentage of the instances of *selja taga*
(back+behind) as a complex postposition that expresses SUPPORT based on the
frequency of the PN lemma.
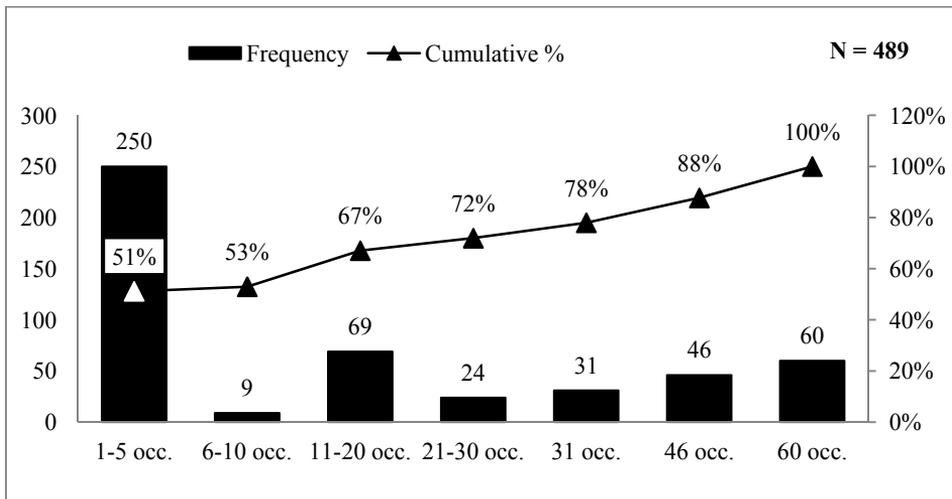
Figure 22. The cumulative percentage of examples of *selja taga* (back+behind) in the
function SUPPORT formed with PN lemmas based on their number of occurrence

Based on the association measures presented in Table 25, the PN lemmas that occurred up to 24 times are not considered strong collocates. As such examples make up 72% of the data, the use of the complex postpositional *selja taga* (back+behind) that expresses SUPPORT can be considered quite productive.

Excluding the examples with proper names reduces the proportion of examples that can be considered to fall under productive use – the amount of examples forms with lemmas that appear up to 5 times is 36%, the amount that occur up to 10 times is 38%, and those that appear up to 30 times is 63%, which leaves the proportion of examples with strong collocates at 37%. Thus, it seems that the productivity is somewhat dependent on the fact that *selja taga* (back+behind) tends to co-occur with proper names.

Figure 23, which gives the cumulative percentage of the examples of SUPPORT based on the frequency of the verb lemma, suggests that the strongest collocates *olema* 'be' and *seisma* 'stand' make up a considerable amount of data. *Olema* 'be' and *seisma* 'stand', which appear in 259 and 119 examples respectively make up 64% of the examples. Thus, the amount of productively formed examples is only 36%. Therefore, based on its use with the verb, the complex postposition *selja taga* (back+behind) that expresses SUPPORT is rather unproductive.
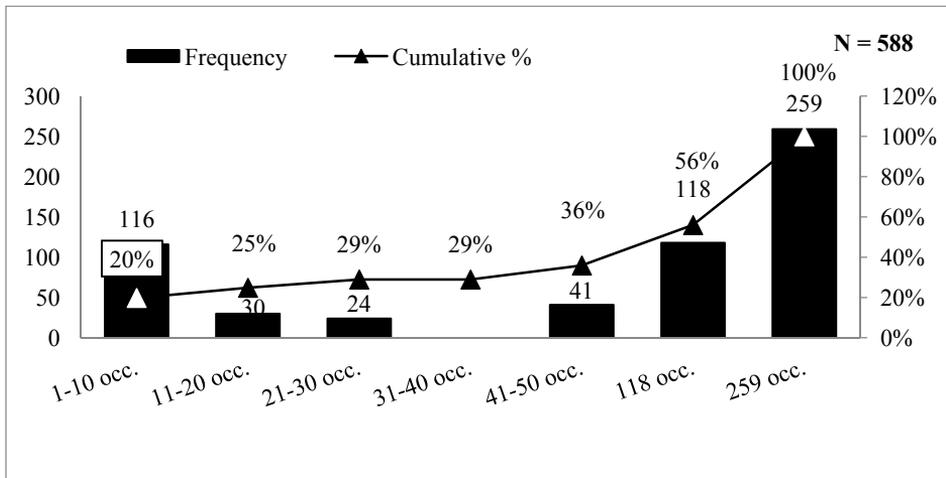
Figure 23. The cumulative percentage of examples of *selja taga* (back+behind) expressing SUPPORT formed with lemmas based on their occurrence

4.6.3.3.3. Summary of the productivity of *selja taga* (back+behind) expressing SUPPORT

When expressing SUPPORT, *selja taga* (back+behind) co-occurs with 202 PN lemmas and 62 verb lemmas. Even though most of these lemmas co-occur infrequently with *selja taga* (back+behind), there are certain frequent lemmas, which proved to be strong collocates of *selja taga* (back+behind) when

expressing SUPPORT. These are pronouns *nemad* 'they', *tema* 's/he' and *enda* 'own' and verbs *seisma* 'stand' and *olema* 'be'. As these lemmas are highly frequent in the context of *selja taga* (back+behind) they are not considered to suggest productive use of the phrase as a complex unit. Based on the analysis of PN lemmas, it was determined that the unproductively formed examples make up approximately one third (28%) of all the examples, which means that close to three fourths (72%) of the examples are formed productively. This indicates that the complex unit *selja taga* (back+behind) is rather productive when expressing SUPPORT. Nevertheless, when the examples with proper names are excluded from the data, the amount of productive uses dropped to 63%. Moreover, the analysis based on the verb lemmas also suggests lower productivity. In this case, the examples with these strongest collocates make up 64% of the data, leaving the amount of productively formed examples at 36%. Therefore, as a complex unit, *selja taga* (back+behind) is rather unproductive when expressing SUPPORT.

4.6.3.4. CONCEALMENT

When expressing CONCEALMENT (*n* = 307), the total number of PN lemmas is 144 and verb lemmas is 77. In this case too, the majority of the lemmas are highly infrequent – 94% of the PN and 90% of the verb lemmas occur up to five times. In this function, proper names make up 14% of the lemmas. However, when the examples with proper names were excluded from the data, the amount of lemmas that occurred up to 5 times was still over 90%. Thus, excluding examples with proper names does not affect the amount of low-frequency lemmas. In the following, analyses of the most frequent PN and verb lemmas of *selja taga* (back+behind) when it expresses CONCEALMENT are presented.

4.6.3.4.1. The strongest collocates of *selja taga* (back+behind) when expressing CONCEALMENT

Table 27 shows the most frequent PN lemmas of *selja taga* (back+behind), their frequency in the etTenTen corpus, the number of times they occur with *selja taga* (back+behind) as a complex postposition, and their log-likelihood score. The lemmas are ranked based on the log-likelihood score. The statistically strongest collocates are separated with the dotted line.

Table 27. The strongest PN collocates of *selja taga* (back+behind) in the function CONCEALMENT

PN lemma	meaning	*n* lemma (in etTenTen)	*n* collocate	log-likelihood score
vanajumal	Great God	366	49	1151
jumal	God	137,099	14	128

PN lemma	meaning	*n* lemma (in etTenTen)	*n* collocate	log-likelihood score
mamma	mama	1,456	5	81
autoriteet	authority	5,174	4	53
nemad	they	178,746	6	41
naine	woman	213,934	6	39
tema	s/he	1,765,208	11	39
skeptik	sceptic	5,020	3	38
üksteise	each other	33,996	4	38
mees	man	303,390	6	35
usa	USA	72,983	4	32
teine	other	717,260	6	25
agent	agent	4,049	2	25
isa	father	78,486	3	21
keegi	someone	267,117	4	21
kes	who	1036,055	6	20
laps	child	475,173	4	17
enda	own	971,567	5	16
riik	state	356,114	3	12

In the case of complex postpositions used to express CONCEALMENT, there is only one highly frequent collocate, i.e. *vanajumal* 'Great God', which appears on 49 occasions. With a log-likelihood score of 1151, this lemma is also statistically the strongest collocate. The absolute frequencies and log-likelihood scores of the rest of the lemmas are considerably lower. Thus, the rest of the PNs are not taken to represent strong collocates. However, *jumal* 'God' – which occurs on 14 occasions – could also be considered as a candidate of a frequent collocate. However, as a single lemma, its log-likelihood value (128) is too low to suggest a strong statistical association with *selja taga* (back+behind), especially compared to *vanajumal* 'Great god'. The reason these two lemmas have been coded separately is that the compound *vanajumal* can be considered to be somewhat lexicalized, and more importantly, the utterance *vanajumala selja taga* (lit. behind Great God's back) forms an idiomatic expression 'secure, protected' (as in (202)–(203)).

(202) *Eesti*   *asu-b*   *tänu*   *oma-ø*   *väiksuse-le*   *ja*
       Estonia   lie-3SG   thanks to   own-GEN   smallness-ALL   and
       isolatsiooni-le   *kui*   ***vanajumala-ø***   ***selja-ø***   ***taga***.
       isolation-ALL   as if   Great God-GEN   back-GEN   behind.LOC
       Lit. Thanks to its small size and isolation, Estonia is situated as if behind the back of God.
       'Thanks to its small size and isolation, Estonia is a safe place.'
       [arvamus.postimees.ee]

(203) *Muidu   ela-b   pensionär nagu* **_vanajumala-ø   selja-ø   taga._**
      else   live-3SG pensioner like   Great God-GEN   back-GEN   behind.LOC
      Lit. Otherwise the pensioner lives as if behind the back of God.
      'Otherwise the pensioner lives comfortably and safely.' [rahvahaal.delfi.ee]

Although the examples with *jumal* 'God' are semantically rather close to the ones that are formed with *vanajumal* 'Great God', they are less fixed. For instance, the idiomatic *vanajumala selja taga* often includes a comparison marker, such as *kui* 'as if' or *nagu* 'like', which is generally not the case with *jumala selja taga* (see example (204)). Moreover, *jumala selja taga* appears in the locative as well as the lative form, whereas *vanajumala selja taga* only appears in the locative.[62] In this respect, the usages with the PN *jumal* 'God' rather resemble that of the other lemmas, which are combined with *selja taga* (back+behind) productively (see examples (205)–(207)).

 (204) *Igasugu   lollikes-te-l   on   suur   vajadus* **_Jumala-ø   selja-ø_**
       all kind   fool-PL-ADE be.3SG big   need   God-GEN   back-GEN
       taha   *peitu-ø   puge-da,   sest   oma-ø   mõistuse-ga*
       behind.LAT   hide-ILL   creep-INF because own-GEN   mind-COM
       na-d   mõel-da   ei   suuda-ø.
       they-PL   think-INF   NEG   able-CONNEG
       Lit. All kinds of fools have a great need to hide behind the back of God because they cannot use their own brains.
       'All kinds of fools have a great need to hide behind God because they cannot use their own brains.' [rahvahaal.delfi.ee]

(205) *Hea   ja   mugav   on   puge-da* **_autoriteeti-de_**
      good   and comfortable be.3SG creep-INF   authority-PL.GEN
      selja-ø   taha   *ja   just   se-da   ma   kavatse-n-gi*
      back-GEN   behind.LAT   and   exactly   this-PRT   I   intend-1SG-CL
      teh-a.
      do-INF
      Lit. It is good and comfortable to hide behind the back of authorities and this exactly what I am planning on doing.
      'It is good and comfortable to hide behind authorities and this exactly what I am planning on doing.' [news.station.ee]

(206) *Üks   osapool seleta-s   avalikult   ära enda-ø   nägemuse-ø,*
      one   party   explain-PST.3SG publicly   off own-GEN   vision-GEN
      teine osapool varju-b   emme-ø   ja **_advokaati-ø seljataha_**.
      other party   hide-3SG mommy-GEN and lawyer-GEN back.behind.LAT
      Lit. One of the parties publicly explained his/her vision, the other party is hiding behind the back of his/her mommy and the lawyers.
      'One of the parties publicly explained his/her vision; the other party is hiding behind his/her mommy and the lawyers.' [www.ekspress.ee]

---

[62]   Although Google search also yields results when searched other forms, the expression clearly prefers the locative form.

(207) *Samas*   *on*   **_karismaatilise-ø_**   **_juhi-ø_**   **_selja-ø_**
furthermore   be.3SG   charismatic-GEN   leader-GEN   back-GEN
taga   *alati*   *väga*   *võimeka-i-d,*   *ent*   *tähelepanu-ta*
behind.LOC   always   very   capable-PL-PRT   but   attention-ABE
jää-nud   *töötaja-i-d* –   *nema-d*   *ole-ks*   *tul-nud*
leave-PST.PTCP   worker-PL-PRT   they-PL   be-COND   must-PST.PTCP
üles leid-a.
up find-INF
Lit. At the same time, behind the back of a charismatic leader there are always
some very capable workers who haven´t received attention – they should have
been found.
'At the same time, behind a charismatic leader there are always some very
capable workers who haven´t received attention – they should have been found.'
[www.director.ee]

There are no clear semantic restrictions to the PN lemmas, in this function but it
can be stated that the PNs can be any noun referring to humans or collectives
(or even more abstract notions) that conceal the TR from responsibility (as in
(205)) and/or threat (as in (206)). Sometimes the CONCEALMENT may be
unintentional or not favourable to the TR (207). However, the CONCEAL-
MENT relationship is usually construed as favourable to the TR, as it is also the
case in the idiomatic expression (*vana*)*jumala selja taga* (lit. behind Great
God's back 'secure, protected').

Table 28 lists the verbs that have the strongest statistical association with
selja taga (back+behind)

Table 28. The strongest verbal collocates of *selja taga* (back+behind) expressing
CONCEALMENT

verb lemma	meaning	*n* lemma in etTenTen	*n* collocate	log-likelihood score
pugema	creep	6,241	75	1,394
peitma/peituma	hide/be hidden	27,196	29	399
olema	be	9,559,606	50	156
elama	live	201,395	16	137
varjuma	shelter	1,299	7	119
varitsema	ambush	1,074	5	83
istuma	sit	68,109	8	75
hoidma/hoiduma	keep	151,324	9	72
tundma	feel	224,072	8	56
juhtima	lead	64,443	5	43
piiluma	peep	4,427	3	39
välja astuma	step out	5,557	3	37
paistma	look	52,167	4	34

verb lemma	meaning	*n* lemma in etTenTen	*n* collocate	log-likelihood score
asuma	lie	118,320	4	27
seisma	stand	75,814	3	21
jääma/jätma	remain/leave	648,139	5	19
plehku pistma	take off	65	1	19
töötama	work	137,400	3	18
niite tõmbama	pull strings	147	1	17

The strongest collocate of *selja taga* (back+behind) expressing CONCEAL-MENT is *pugema* 'creep'. *Pugema* co-occurs with *selja taga* (back+behind) 75 times and has the log-likelihood score of 1394, and, as such, is the strongest collocate. The rest of the lemmas are considered to have a weaker association with *selja taga* (back+behind) because the log-likelihood values are considerably lower (399 and less). However, as most of the lemmas occur on less than 10 occasions, a few other frequent lemmas stand out in the data. One of them is *peitma/peituma* 'hide/be hidden', which is semantically close to the most common lemma *pugema* 'creep' (see examples (208) and (209)). There are other similar verbs among the strongest collocates of *selja taga* (back+behind), e.g. *varjuma* 'shelter'. It was discussed above that in this function, *selja taga* (back+behind) often co-occurs with the PN *vanajumal* 'Great God' or *jumal* 'God'. It was observed that the verb *elama* 'live', which does not have a strong correlation with *selja taga* (back+behind) but which is still one of the most frequent verbs in this function (16 occurrences), is also often associated with such usages (see example (210)). In this respect, such usages are also considered to belong to the group of unproductively formed examples. In the following, I present the amount of examples that are formed productively.

(208) *... ja   püüd-es   enda-l   hinge-ø   sääst-a   on   muidugi*
       and   try-GER   own-ADE   soul-PRT   save-INF   be.3SG   of course
       *lahendus   **puge-da**   jänki-de   **seljataha**   ja*
       solution   creep-INF   yankee-PL.GEN   back.behind.LAT   and
       siuna-ta   iga-l   võimaluse-l   Venemaa-d.
       beshrew-INF   every-ADE   opportunity-ADE   Russia-PRT
       Lit. … and trying to save one's soul it is of course a solution to creep behind the back of the Yankees and beshrew Russia at every opportunity.
       '… and trying to save one's soul it is of course a solution to hide behind the Yankees and beshrew Russia at every opportunity.' [www.ekspress.ee]

(209) *Vana-l   don   Corleone-l   ol-i   ainult   mõnevõrra*
       old-ADE   don   Corleone-ADE   be-PST.3SG   only   somewhat
       rohkem   positiivse-i-d   omadus-i ,   näiteks   ei
       more   positive-PL-PRT   quality-PL.PRT   for example   NEG
       peitu-nud**   ta   kogu   aeg   nais-te   **seljataha
   hide-PST.PTCP   s/he   all   time   woman-PL.GEN   back.behind.LA

kui   kitsas   käe-s.
if   narrow   hand-INE
Lit. The old don Corleone had only a few more positive qualities; for example,
he did not constantly hide behind the backs of women when times were tough.
'The old don Corleone had only a few more positive qualities; for example, he
did not constantly hide behind women when times were tough.' [www.epl.ee]

(210) *Ne-i-s   valitse-s   mingi   loomulik   rahu –   otsekui*
      they-PL-INE   dominate-PST.3SG   some   natural   calm   as if
      ela-ta-ks   *tõepoolest   Jumala-ø   **selja**-ø   **taga***.
      live-IMPS-COND   really   God-GEN   back-GEN   behind.LOC
      Lit. There was some kind of a natural peace in them – as if they really lived
      behind the back of God.
      'There was some kind of a natural peace in them – as if they really lived
      protected by God.' [reisiajakiri.gomaailm.ee]

4.6.3.4.2. The productive use of *selja taga* (back+behind) expressing CONCEALMENT

As the use of *selja taga* in the idiomatic expression clearly does not constitute a productive use of *selja taga* (back+behind), in this case, too, the amount of examples productively formed examples is determined. Figure 24 shows the cumulative percentage of the instances of *selja taga* (back+behind) as a complex unit that expresses CONCEALMENT based on the frequency of the PN lemma.
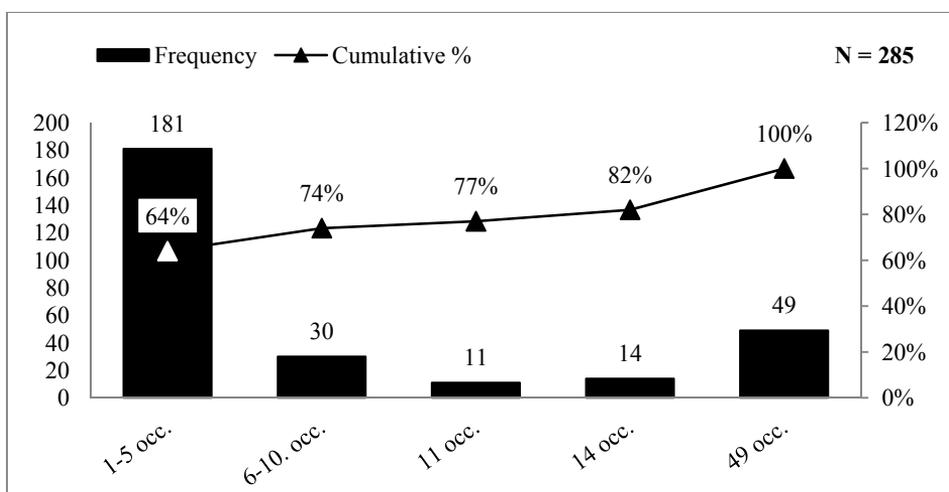
Figure 24. The cumulative percentage of examples of *selja taga* (back+behind) in the function CONCEALMENT formed with PN lemmas based on their number of occurrence

167

The data show that about three-quarters of the examples are productively formed. It can be observed in Figure 24 that 74% (211 examples out of 285) are formed with PNs that occur up to 5 times, and 74% (211 examples out of 285) occur up to 10 times. Based on their low absolute frequencies and log-likelihood scores (see Table 27) lemmas that occurred up to 10 with *selja taga* (back+behind) are considered to account for the productive use of the complex postposition. Even though *tema* 's/he' occurred on 11 occasions, based on its log-likelihood value (39) it is considered not to belong with the strong collocates. Thus, the total amount of productively formed examples is 77%. If the lemma *jumal* 'God' is considered to belong with strong collocates, the amount of non-productive examples is 23%, if the non-productive examples are considered to include the utterances *vanajumala selja taga* (lit. behind Great god's back 'secure, protected') only, the amount of non-productively formed examples is 18%. When the examples with proper names are excluded from the data, the amount of productively formed examples remains about the same (82%).

The amount of productively formed examples based on the analysis of the verb lemma is shown in Figure 25. As mentioned above, in the case of CONCEALMENT, there is one extremely strong verbal collocate of *selja taga* (back+behind) – *pugema* 'creep', which occurs in the data 75 times. Such usages make up 25% of the examples. Therefore, the amount of productively formed examples would be 75%. However, if examples with other more frequent lemmas *peitma/peituma* 'hide/be hidden', *olema* 'be', and *elama* 'live' also considered not to amount to productive use of the complex postposition, the amount of productively formed examples would be 43%.
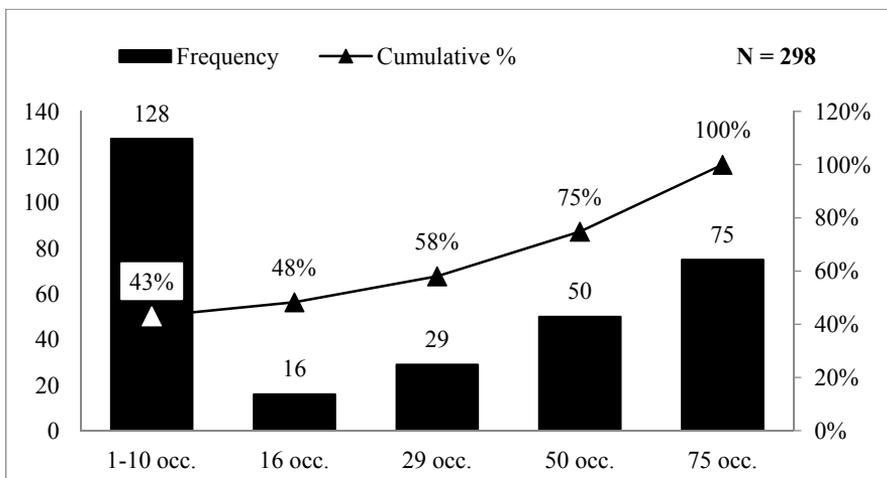
Figure 25. The cumulative percentage of examples of *selja taga* (back+behind) expressing CONCEALMENT formed with verb lemmas based on their occurrence

4.6.3.4.3. Summary of the productivity of *selja taga* (back+behind) when expressing COCEALMENT

When used to express CONCEALMENT *selja taga* (back+behind) takes 144 PN lemmas and 77 verb lemmas. Most of these lemmas are rather infrequent in the company of *selja taga* (back+behind) – over 90% of both PN and verb lemmas occur up to five times.

Based on the analysis of PN lemmas, the complex unit *selja taga* (back+behind) that expresses CONCEALMENT has one very strong collocate – *vanajumal* 'Great God', with which *selja taga* forms an idiomatic expression (*nagu*) *vanajumala selja taga* 'secure, protected'. The data show that such usages make up 18% of the uses. Thus, despite of the fact that some of the uses of *selja taga* involve idioms, most (82%) of the data sample represents uses where the complex postposition is used productively.

However, the analysis of verb lemmas, there is one extremely frequent lemma that also has a strong statistical association with *selja taga* – *pugema* 'creep'. Examples that include this lemma make up 25% of the data, thus leaving 75% of the examples to represent productive use. However, as most of the verb lemmas are quite infrequent (less than 10 occurrences), additionally, verbs *peitma/peituma* 'hide/be hidden', *olema* 'be', and *elama* 'live' stood out as more frequent and strongly associated with the phrase. The latter is also associated with the examples that were considered as idiomatic expressions. Thus, if such examples are also excluded from the productively formed examples, only 43% of examples account for the productive use of *selja taga* (back+behind). In this case, the complex function word *selja taga* (back+behind) expressing CONCEALMENT can be considered rather non-productive.

4.6.4. *Käe kõrval* (hand+beside)

Although *käe kõrval* (hand+beside) mostly occurs as a complex unit (see sections 4.2), the absolute number of such examples remains quite small (*n* = 771). Within these 771 examples, *käe kõrval* (hand+beside) takes 43 PN and 95 verb lemmas.[63] In this case, too, most of the lemmas occur only on a few occasions. The data suggests that 88% of PN lemmas and 83% of PN lemmas occur up to five times. The data suggests that the complex unit *käe kõrval* (hand+beside) has 2 frequent verbal collocates and 5 frequent PN lemmas, which make up 11% of the lemmas. This is considerably larger amount than in case of other phrases. The frequent lemmas and their collocational strength are discussed in the following section.

---

[63] The number of verb lemmas is more than two times higher than that of the PN lemmas because *käe kõrval* (hand+beside) is mostly (85%) used as a complex adverb (see section 4.4.).

4.6.4.1. The strongest collocates of *käe kõrval* (hand+beside) as a complex unit

Table 29 shows that he strongest PN collocate of the complex postpositional *käe kõrval* (hand+beside) is without doubt *ema* 'mother' (log-likelihood score 407). With a total of 26 occurrences, it is also the most frequently occurring PN. Based on absolute frequencies and the log-likelihood score, the lemmas *isa* 'father' and *vanaema* 'grandmother' are also considered to be strongly associated with the complex postpositional *käe kõrval* (hand+beside). The pronouns *oma* 'own' and *tema* 's/he' – both of which occurred 10 times – have a weaker association (log-likelihood scores 106 and 101 respectively) but are still considered to belong to the group of strong collocates. The other lemmas only occur a few times and are not considered to form a fixed expression with the complex postpositional *käe kõrval* (hand+beside).

Table 29. The strongest PN collocates of *käe kõrval* (hand + beside) as a complex unit

PN lemma	meaning	*n* (lemma in etTenTen)	*n* collocate	log-likelihood score
ema	mother	107,212	26	407
isa	father	78,486	13	212
vanaema	grandmother	17,354	8	154
oma	own	1,346,491	10	106
tema	s/he	1,765,208	10	101
emme	mommy	7,678	3	63
vanem	parent	110,813	4	62
naine	woman	213,934	4	57
saatja	escort	4,979	2	44
vanaisa	grandfather	8,607	2	41
Perkmann	Perkmann*	15	1	33
Kirkegaard	Kirkegaard*	23	1	33
õpetaja	teacher	125,700	2	31
jumal	god	137,099	2	30
nemad	they	178,746	2	29
kaugelviibija	absentee	707	1	26
Potter	Potter*	737	1	26
Helmi	Helmi*	798	1	25
koondislane	member of the national team	974	1	25
Leivo	Leivo*	1,469	1	24

* Proper name

The strongest collocates are not considered to form an idiomatic expression with the complex postpositional *käe kõrval* (hand+beside), but are rather examples of the most typical use. Thus, examples such as (211) express the most basic meaning of the complex function word *käe kõrval* (hand+beside), i.e. refer to relationship between a mother and a child that includes the physical aspect (next to one's hand, holding hands) as well as the component of accompaniment. Such meaning is also carried by the examples with the lemmas *isa* 'father', *vanaema* 'grandmother', *vanem* 'parent', *vanaisa* 'grandfather', and especially *emme* 'mommy' (see Table 29). Thus, to a certain extent, it seems that the PNs of *käe kõrval* (hand+beside) are also thematically restricted. However, the use of *käe kõrval* (hand+beside) is not confined to PNs referring to family members. It can also occur with other authority figures, such as *õpetaja* 'teacher'. Such usages are illustrated by (212), where the sense of physical adjacency is lost or at least marginal and the sense of (guiding) accompaniment more salient. There are also examples of purely abstract uses (see example (213)), where the physical component has entirely disappeared. However, these are not particularly frequent in the data (see section 4.3.4). Usually, *külje all* (side+under) contains both – physical as well as abstract accompaniment.

(211) *Maarja   on   balleti-ga   tegele-nud   alates   seitsmenda-st*
       Maarja   be.3SG   ballet-COM   exercise-PST.PTCP   since   seventh-ELA
       *eluaasta-st,   kui   sead-is   **ema-ø**   **käe-ø**   **kõrval***
       year-ELA   when   set-PST.3SG   mother-GEN   hand-GEN   behind.LOC
       sammu-d   Pärnu-ø   Kunsti-de   Maja-ø   balletiringi-ø.
       step-PL   Pärnu-GEN   Art-PL.GEN   House-GEN   ballet school-ILL
       Lit. Maarja has been doing ballet since she was seven, when she went to the ballet school in Pärnu Kunstide Maja beside the hand of her mother.
       'Maarja has been doing ballet since she was seven, when she went to the ballet school in Pärnu Kunstide Maja alongside her mother.' [tartutants.ee]

(212) *Üks   esimes-i   on   kooliraamatukogu,   kuhu*
       one   first-PL.PRT   be.3SG   school library   where
       *algklassiõpilas-i   **õpetaja-ø**   **käekõrval***
       elementary school student-PL.PRT   teacher-GEN   hand.beside.LOC
       vii-a-kse   ja   räägi-ta-kse,   et   enne   raamatu-ø
       bring-IMPS-PRS   and   talk-IMPS-PRS   that before book-GEN
       kättevõtmis-t   tule-b   käe-d   puhta-ks   pes-ta.
       take in hand-PRT   must-3SG   hand-PL   clean-TRL   wash-INF
       Lit. One of the firsts is the school library, where elementary school students are taken beside the hand of their teacher and where they are told that before they pick up a book, they have to wash their hands.
       'One of the firsts is the school library, where elementary school students are taken alongside their teacher and where they are told that before they pick up a book, they have to wash their hands.' [www.virumaateataja.ee]

(213) **_Kierkegaardi-Ø_**   **_käekõrval_**   _tee-n_   _ma_   _siin_   _katse-t_
    Kieregaard-GEN   hand.beside.LOC   make-1SG I   here   try-PRT
    minn-a   _teis-t_   _tee-d._
    go-INF   another-PRT   way-PRT
    Lit. Beside the hand of Kierkegaard I am trying to go another way.
    '(Together) with Kierkegaard I am trying to go another way' [www.eestikirik.ee]

The strogest verbal collocates are given in Table 30. It can be observed in that _lükkama_ 'push' co-occurs with _käe kõrval_ (hand+beside) 111 times and has the log-likelihood score of 1,585, and, as such, is the strongest collocate. The lemma _võtma_ 'take', which occurred in 128 instances, is a close second with the score 1,125. The rest of the verb lemmas are considerably less frequent (62 and fewer occurrences) and received considerably lower scores (611 and lower).

Table 30. The strongest verbal collocates of _käe kõrval_ (hand+beside) as a complex unit

verb lemma	meaning	_n_ lemma in etTenTen	_n_ collocate	log-likelihood score
lükkama	push	30,878	111	1,585
võtma	take	552,338	128	1,125
talutama	lead	570	31	611
viima	take (somewhere)	183,356	62	592
kõndima/kõnnitama	walk (/someone)	17,731	28	354
jalutama	stroll	19,280	26	320
käima	go	302,976	29	204
tooma	bring	229,994	16	102
minema	go	605,794	17	78
vedama	pull	31,615	8	72
astuma	step	64,581	9	70
jooksma/jooksutama	run (/someone)	47,144	8	65
juhatama	lead	14,329	6	60
saatma	send	112,953	8	51
sammuma	stride	3,670	4	48
sõitma/sõidutama	drive (/someone)	132,485	6	33
tatsama	toddle	678	2	28
õpetama	teach	50,093	4	27
hoidma	hold	143,450	5	25
taterdama	pattle	2	1	24

Lükkama 'push' occurs mostly in sentences where _käe all_ (hand+under) is used to express animate LM's position relative to an inanimate TR, which is mostly a bicycle. Such usages are illustrated in (214). _Võtma_ 'take', on the other hand, occurs in sentences with animate TRs, who are mostly children (as in (215)) but

also adults who need some sort of guidance. This is in accordance with the results of the above analysis, which suggests that PNs of *käe kõrval* (hand+beside) are typically authority figures (e.g. parents, teachers). It can be observed in Table 30 that most of the verbs among the 30 strongest collocates of *käe kõrval* (hand+beside) as a complex unit are verbs of movement (e.g. *viima* 'take (someone to somewhere)' (as in (216), *minema* 'go', *astuma* 'step', *jooksma* 'run'). Thus the typical instance of *käe kõrval* (hand+beside) expresses the relation of a helpless or inanimate entity in movement with an (authoritative) animate LM. The examples with the strongest collocates illustrated in (214) and (215) are not considered to be idioms but they are also not considered to represent productively formed examples. The amount of productively formed examples is determined in the following section.

(214) **Lükka-si-n**   *ratas-t*   **käe-ø**   **kõrval**   *sama-ø*
push-PST-1SG   bicycle-PRT   hand-GEN   beside.LOC   same-PRT
tee-d   *pidi*   *tagasi*   *Tartu-sse.*
road-PRT   along   back to   Tartu-ILL
Lit. I pushed the bike beside my hand back to Tartu along the same road.
'I pushed the bike beside me back to Tartu along the same road.'
[www.lounaleht.ee]

(215) *ol-i-n*   *ju*   *tite-st*   *saati*   *käi-nud*   *Sõduri-ø*
be-PST-1SG   well   baby-ELA   since   go-PST.PTCP   Soldier-GEN
juurde   *lill-i*   *vii-ma-s,*   *ol-i-n*   *õnnelik*
to   flower-PL.PRT   take-SUP-INE   be-PST-1SG   happy
ol-nud,   *nee-d*   *ol-i-d*   *alati*   *ilusa-d*   *päeva-d,*
be-PST.PTCP   this-PL   be-PST-3PL   always   beautiful-PL   day-PL
papa   **võtt-is**   *min-d*   **käe-ø**   **kõrvale**   *ja*   *mina*
daddy   take-PST.3SG   I-PRT   hand-GEN   beside.LAT   and   I
sa-i-n   *kuhja-ø*   *nelk-e*   *kätte-ø…*
get-PST-1SG   pile-GEN   carnation-PL.PRT   hand-ILL
Lit. I had gone to the Soldier since I was a baby, I has been happy because these were always beautiful days, daddy took me beside his hand and I had bunch of carnations in my hand.
'I had gone to the Soldier since I was a baby, I has been happy because these were always beautiful days, daddy took me beside him andI had bunch of carnations in my hand.' [www.poogen.ee]

(216) *"Nupp*   *tööle!"*   *project*   **vii-b**   *noore-d*   **käekõrval**
noddle   to work   project   take-3SG   young people-PL   hand.beside.LOC
tööturu-le.
job market-ALL
Lit. The project "Nupp tööle!" takes young people beside the hand to the job market.
'The project "Nupp tööle!" takes young people to the job market.'
[www.nupptoole.ee]

173

### 4.6.4.2. The productive use of *käe kõrval* (hand+beside) as complex unit

Figure 26 shows the cumulative percentage of the instances of *käe kõrval* (hand+beside) as a complex postposition based on the frequency of the PN lemma.
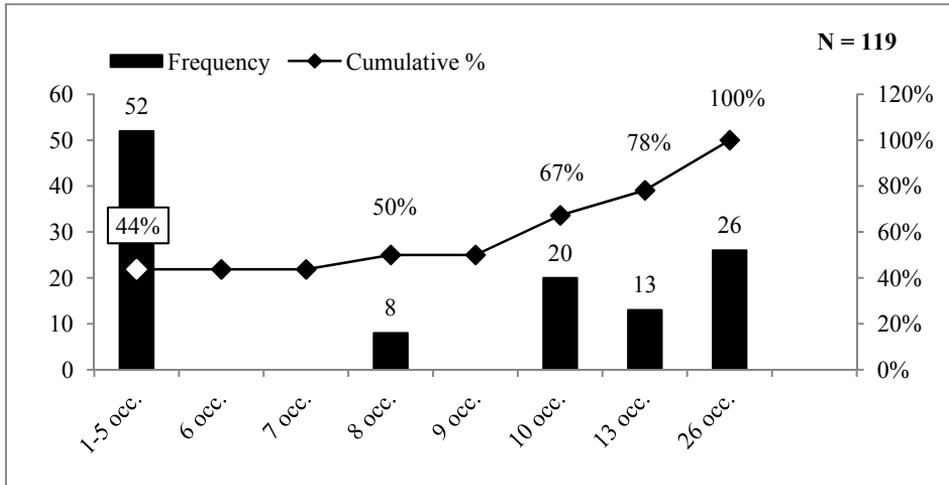
Figure 26. The cumulative percentage of examples of *käe kõrval* (hand+beside) formed with PN lemmas based on their number of occurrence

It can be observed in Figure 26 that 44% of the examples (52 out of 119) are formed with the PN lemmas that appear 1–5 times. As the data sample is relatively small (119 examples with PN present), all of the more frequent lemmas had a stronger statistical association with the complex postpositional *käe kõrval* (hand+beside) and are therefore not considered to suggest productive use. Thus, over half (56%) of the examples are formed with strong collocates, which indicates that the complex postpositional *käe kõrval* (hand+beside) is considerably less productive than the other phrases analyzed above.

The amount of productively formed examples based on the analysis of the verb lemma is shown in Figure 27. The data show that the strongest collocates – *lükkama* 'push' and *võtma* 'take', which occur on 111 and 128 occasions respectively make up 37% of all the examples where *käe kõrval* (hand+beside) is used as a complex unit. Therefore, 63% of the examples are formed with weaker collocates of *käe kõrval* (hand+beside). Thus, the amount of examples that are formed productively is 63%, which indicates the same grammatical status as the more frequent phrases (*käe all* (hand+under) and *külje all* (side+under)).
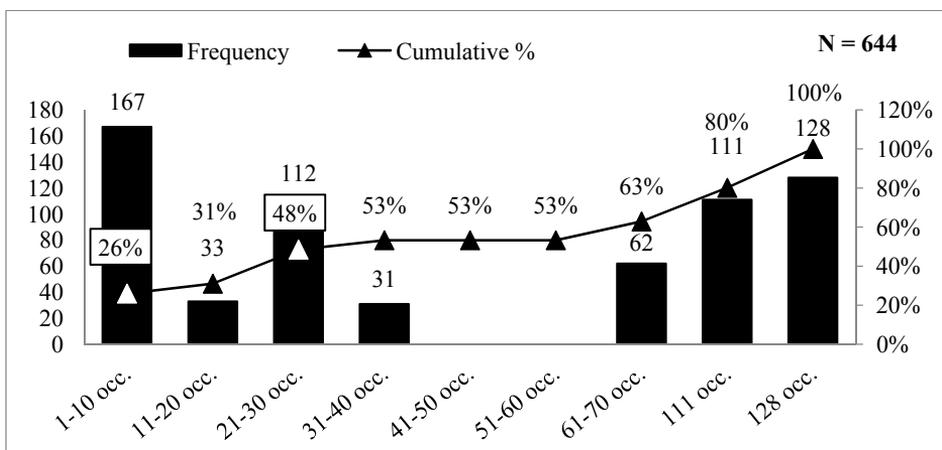
Figure 27. The cumulative percentage of examples of *käe kõrval* (hand+beside) formed with lemmas based on their occurrence

4.6.4.3. Summary of the productivity of *käe kõrval* (hand+under) as a complex item

As a complex unit, *käe kõrval* (hand+beside) co-occurs with 43 PN and 95 verb lemmas. Around 90% of the lemmas occur a few times (up to five occasions) but as in every case analyzed so far, there are a few lemmas that are strongly associated with *käe kõrval* (hand+beside).

The examples that are formed with the strong PN collocates – *ema* 'mother', *isa* 'father', *vanaema* 'grandmother', *tema* 's/he', and *oma* 'own' – make up over half of all the examples of the complex postpositional *käe kõrval* (hand+beside). Therefore, based on the analysis of the PN lemmas, the productivity of the complex unit *käe kõrval* (hand+beside) is considerably lower than in the cases of other studied phrases discussed above. However, the analysis of the verb lemma suggest a different conclusion. As there are only two strong collocates – *lükkama* 'push' and *võtma* 'take – , which appear in 37% of the examples. Thus, the analysis suggests that 63% of the examples where *käe kõrval* (hand+beside) is used as a complex function word are formed productively.

4.6.5. *Kaela peal* (neck+on)

Compared to the other studied phrases, *kaela peal* (neck+on) is not a particularly frequent phrase – it is realized as complex unit on 87 occasions (see section 4.2). The data show that as a complex unit, *kaela peal* (neck+on) takes 25 PN lemmas and 22 verb lemmas. As was also observed in the case of other phrases, most of the lemmas are used infrequently – 18 out of 25 PN lemmas (72%) occur up to two times and 82% (18 out of 22) of the verb lemmas occur up to five times. As the number of examples of *kaela peal* (neckt+on) is limited, there is no clearly

distinguished group of highly frequent lemmas. However, it can be observed that there are PN and verb lemmas that occur more frequently than others.

4.6.5.1. The strongest collocates of *kaela peal* (neck+on) as a complex unit

The strongest PN collocates that occur with the complex postpositional *kaela peal* (neck+on) are presented in Table 31. The data show that the most frequently occurring PN is *mina* 'I' with six occurrences. However, as the first person pronoun is a very frequent lemma in general, it is the lemma *vanem* 'parent' (5 occurrences) that is statistically the strongest collocate (log-likelihood value 55) of complex postpositional *kaela peal* (neck+peal). Moreover, *ema* 'mother', *nemad* 'they', *keegi* 'someone', and *oma* 'own' can be also considered as strongly associated with the complex postpositional *kaela peal* (neck+on).

Table 31. The strongest PN collocates of *kaela peal* (neck+on)as a complex unit

PN lemma	meaning	*n* lemma (in etTenTen)	*n* collocate	log-likelihood score
vanem	parent	110,813	5	55
mina	I	1,865,803	6	34
ema	mother	107,212	3	30
nemad	they	1,758,591	5	27
keegi	someone	267,117	3	25
oma	own	1,346,491	4	22
rahvas	people	141,245	2	17
enda	own	971,567	3	17
naine	woman	213,934	2	16
alluv	employee	5,291	1	14
EV	Republic of Estonia*	6,143	1	14
tädi	aunt	8,155	1	13
laps	kid	475,173	2	12
maksumaksja	taxpayer	14,755	1	12
lähedane	close one	20,621	1	11
õde	sister	24,295	1	11
teine	other	718,828	2	11
omanik	owner	58,687	1	9
USA	USA*	72,983	1	9
süsteem	system	75,426	1	9

* Proper name

The utterance *vanemate kaela peal* 'burden to one's parents' is clearly the most prototypical use of the phrase, which is most commonly used to refer to the burden that one can put/lay on their relatives (see example (217)). Accordingly, there are other thematically close lemmas that are used together with *kaela peal* (neck+on), such as the fairly strongly associated *ema* 'mother' (see example (218)), but also the less frequent *õde* 'sister', *lähedane* 'close one', *laps* 'kid', *tädi* 'aunt/old lady', *naine* 'woman/wife' etc. (see Table 31). In addition to *mina* 'I', there are three more pronouns among the top 20 lemmas – *nemad* 'they', *keegi* 'someone', *oma* 'own'. These may be also used to refer to family members as in example (219), where pronoun *nemad* 'they' is used to refer to one's parents.

(217) *... kelle-l   raha-ø   p-ole   piisavalt,   et   en-d   ära*
       who-ADE   money-PRT   NEG-be   enough   that   own-PRT   off
       *elata-da,   see   jää-b-ki   oma-ø   **vanema-te**   **kaelapeale***
       manage-INF   this   stay-3SG-CL   own-GEN   parent-PL.GEN   neck.on.LAT
       Lit. Those who do not have enough money to manage on their own, remain on
       the necks of their parents.
       'Those who do not have enough money to manage on their own, remain
       dependent on their parents.' [rahvahaal.delfi.ee]

(218) *Ta   ol-i   kuusteist   aasta-t   vana,   kui   otsusta-s,*
       s/he   be-PST.3SG   sixteen   year-PRT   old   when   decide-PST.3SG
       *et   aitab   **ema-ø**   **kaela-ø**   **peal**   elamise-st,   ja*
       that   enough   mother-GEN   neck-GEN   on.LOC   living-ELA   and
       koli-s   Tallinna-ø.
       move-PST.3SG   Tallinn-ILL
       Lit. S/he was sixteen when s/he decided it was enough for him/her to be living
       on the neck of his/her mother and s/he moved to Tallinn.
       'S/he was sixteen when s/he decided it was enough for him/her being dependent
       on his/her mother and s/he moved to Tallinn.' [www.ekspress.ee]

(219) *... kui   te-i-d   ema-isa   on   sellis-te-ks*
       if   you-PL-PRT   mom-dad   be.3PL   this kind-PL-TRL
       *kasvata-nud,   siis   ol-ge   aga   **nen-de**   **kaela-ø**   **peal**.*
       raise-PST.PTCP   then   be-IMP   but   they-PL.GEN   neck-GEN   on.LOC
       Lit. If you mother and father have raised you to be like this, keep on being on
       their necks.
       'If you mother and father have raised you to be like this, keep on being
       dependent on them.' [noortehaal.delfi.ee]

The use of *kaela peal* (neck+on) is not confined to express the burden of one's family members. It can also be used to refer to other types of similar relations. For instance, when the complex postpositional *kaela peal* (neck+on) is used with lemmas such as *alluv* 'employee' or *maksumaksja* 'taxpayer' (see example (220)) or even countries (Republic of Estonia and the USA). Thus, similarly to

the phrases *käe all* (hand+under), *külje all* (side+under), *selja taga* (back+behind) when expressing SUPPORT, and *käe kõrval* (hand+beside), it seems that the complex postpositional *kaela peal* (neck+on) also has a preference of PNs of a certain theme, but it is not confined to such uses.

(220) *RK,*   *valitsus*   *ja*   *kogu*   **<u>maksumaksja-ø</u>**   **kaela-ø**
     parliament,   government   and   whole   taxpayer-GEN   neck-GEN
     peal   *ela-v*   *JUHTIV*   *ametnikkond*   *saa-vad*   *küll*.
     on.LOC   live-PTCP   controlling   public cervant   can-3PL   indeed
     Lit. The leading public officials living on the neck of the Parliament, Government and taxpayers can.
     'The leading public officials dependent on the Parliament, Government and taxpayers can.' [www.delfi.ee]

It can be observed in Table 32 that the strongest collocates of *kaela peal* (neck+on) as a complex unit are *istuma* 'sit', *jääma/jätma* 'remain/leave', *olema* 'be', and *elama* 'live'. *Istuma* 'sit' co-occurs with the complex *kaela peal* (neck+on) 10, *jääma/jätma* 'remain/leave' 13, *olema* 'be' 23, and *elama* 'live' 9 times. The absolute frequencies of these lemmas are rather low but as there are only 87 examples, they make up a considerable amount of data. The rest of the lemmas occur on 5 occasions or less.

Table 32. The strongest verbal collocates of *kaela peal* (neck+on) as a complex unit

verb lemma	meaning	*n lemma* in etTenTen	*n* collocate	log-likelihood score
istuma	sit	68,109	10	122
jääma/jätma	remain/leave	648,139	13	107
olema	be	9,559,606	23	92
elama	live	201,395	9	89
võtma	take	552,338	5	33
veeretama	roll	2,991	2	31
lulli lööma	horse around	72	1	21
agiteerima	agitate	194	1	19
parasiteerima	parasite	280	1	19
upitama	push (up)	1,694	1	15
pagema	escape	1,878	1	15
vinguma	whine	4,024	1	13
passima	wait	4,112	1	13
maanduma	land	5,215	1	13
rippuma	hang	6,438	1	12
tahtma	want	361,620	2	11
ronima	climb	15,605	1	11
saama	get	1,871,918	3	10

verb lemma	meaning	*n lemma* in etTenTen	*n* collocate	log-likelihood score
soovitama	recommend	83,992	1	7
hoidma	hold	143,450	1	6
tooma	bring	229,994	1	5
käima	go	302,976	1	5
panama	put	347,171	1	4

Typical usage of examples with *istuma* 'sit' is exemplified in (221). Because of the figurative use of the verb, such examples are considered as idiomatic expressions. The rest of the uses with strong verbal collocates (see examples (222)–(223)) are more neutral but, nevertheless, because of their frequent collocation with *kaela peal* (neck+on), they are not considered to represent productive use of the phrase, the amount of which is determined in the following section.

(221) **Istu-b-ki** siis selline ülekasvanud mehike, kelle-l juba
sit-3SG-CL then this kind overgrown man who-ADE already
hamba-i-d ja juukse-i-d vähevõitu, ning illusiooni-d ka
tooth-PL-PRT and hair-PL-PRT little and illusion-PL also
ammu läi-nud, süüdimatult oma-ø ema-ø **kaela-ø**
long go-PST.PTCP irresponsibly own-GEN mother-GEN neck-GEN
peal.
on.LOC
Lit. There it is, an overgrown man, who has few teeth and hair and whose illusions are long gone, irresponsibly sitting on his mother neck.
'There it is, an overgrown man, who has few teeth and hair and whose illusions are long gone, irresponsibly dependent on his mother.' [www.24tundi.ee]

(222) *Aga mina pea-n arvatavasti oma-ø las-te* **kaela-ø**
but I have to-1SG probably own-GEN kid-PL.GEN neck-GEN
peale jää-ma *mis mu-lle kuidagi ei meeldi-ø*.
on.LAT leave-SUP what I-ALL someway NEG like-CONNEG
Lit. But I have to most probably remain on the neck of my children and I do not like it a bit.
'But I have to most probably be dependent on my children and I do not like it a bit.' [www.ekspress.ee]

(223) *Jah   mis   teh-a,   keegi   ei   taha-ø   ju   **võt-ta***
yes   what   do-INF   nobody   NEG   want-CONNEG   well   take-INF
*oma-ø   **kaelapeale**   vastutus-t,   sellepärast   ka*
own-GEN   neck.on.LAT   responsibility-PRT   because   also
lihtsa-ma-ø   vastupanu-ø   tee-d   min-dud   ja   nagu
easy-COMP.GEN   resistance-GEN   way-PRT   go-PST.PTCP   and   like
aru-ø   saa-da   siis   ka   korralduskulu-d   väikse-ma-d,
mind-PRT   get-INF   then   also   managing cost-PL   small-COMP-PL
nn   masu-ø   aeg.
so-called   economic depression-GEN   time

Lit. Yes, what can you do, no one wants to take the responsibility on their neck, that´s why we have taken the easier way out and as you can see, the managing costs have been smaller, it is the so called recession after all.

'Yes, what can you do, no one wants to take on the responsibility, that´s why we have taken the easier way out and as you can see, the managing costs have been smaller, it is the so called recession after all.' [foorum.mosseliiga.ee]

4.6.5.2. The productive use of *kaela peal* (neck+on) as complex unit

Since there is very little data on the complex postpositional *kaela peal* (neck+on), it is difficult to unambiguously differentiate between strong and weak collocates. Since one of the lemmas that occurred three times (*ema* 'mother') belongs to the strongest collocates of *kaela peal* (neck+on), only the lemmas that occur once or twice are considered to represent productive uses. Figure 28 shows that such lemmas only account for 43% of the examples, i.e. over half of the data (57%) are examples of more or less fixed expressions. Thus, it is concluded that based on the analysis of PN lemmas, the complex postpositional *kaela peal* (neck+on) is similar to *käe kõrval* (hand+beside) in being considerably less productive than the other studied phrases.
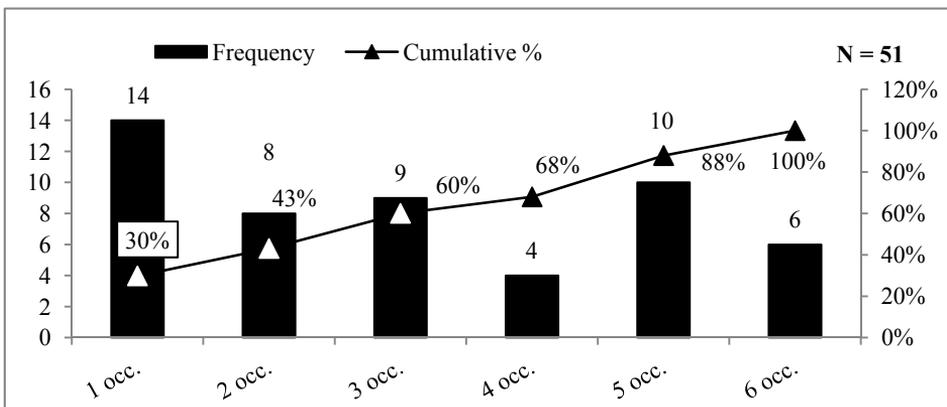
Figure 28. The cumulative percentage of examples of *kaela peal* (neck+on) formed with PN lemmas based on the number of occurrence

The amount of productively formed examples based on the analysis of the verb lemma can be observed in Figure 29, which gives the cumulative percentage of the instances of *kaela peal* (neck+on) as a complex unit.
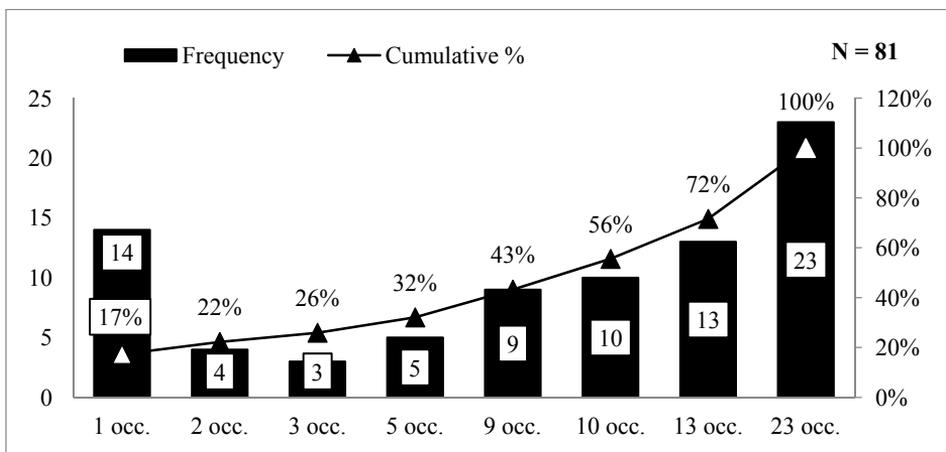
Figure 29. The cumulative percentage of examples of *kaela peal* (neck+on) formed with lemmas based on their occurrence

Figure 29 shows that the strongest collocates of *kaela peal* (neck+on) as a complex unit – *istuma* 'sit', *jääma/jätma* 'remain/leave', *olema* 'be', and *elama* 'live' – which occur on 10, 13, 23, and 9 occasions respectively, make up over two thirds (68%) of the usages of *kaela peal* (neck+on) as a complex unit. This suggest that 32% of the examples account for productive use of *kaela peal* (neck+on), which does not suggest productivity and is, hence, also indicative of lower grammatical status.

4.6.5.3. Summary of the productivity of *kaela peal* (neck+on) as complex unit

Kaela peal (neck+on) is seldom used as a complex postposition, and it is the least productive of the phrases analyzed so far. Within the 87 examples where *kaela peal* (neck+on) functions as a complex unit, it was modified by 25 PN and 22 verb lemmas. Even though most of the lemmas occur on a few occasions, *kaela peal* (neck+on) has strong collocates among the PN as well as the verbs. Even though all of such usages are not idiomatic, they are not considered to represent productive use of the complex item.

Among the PNs, the strong collocates make up over half (57%) of all the usages of *kaela peal* (neck+on). Thus, regarding the PNs, the complex postpositional *kaela peal* (neck+on) is not considered very productive. There are four verb lemmas that have a strong statistical association with *kaela peal* (neck+on) – *istuma* 'sit', *jääma/jätma* 'remain/leave', *olema* 'be', and *elama* 'live'. Such examples make up 68% of the data. Thus, as can be expected based

on the low frequency of *kaela peal* (neck+under) it is not considered as productive as the more frequent phrases analyzed above .

4.6.6. *Jalge all* (feet+under)

As mentioned in section 4.3.6, as a complex unit *jalge all feet* (feet+under) expresses two functions – OPPRESSION and INCEPTIVENESS. As these functions have developed via different grammaticalization paths, they must be analyzed separately.

4.6.6.1. INCEPTIVENESS

It was shown in section 4.4 that *jalge all* (feet+under) occurs as a complex unit expressing INCEPTIVENESS on 344 occasions. Within these 344 examples, *jalge all* (feet+under) occurs with just 5 PN lemmas and 7 verb lemmas. As the data suggests that the examples where *jalge all* (feet+under) expresses INCEPTIVENESS ($n = 5$) are each formed with a different PN lemma, these will not be discussed any further.[64] It can be observed in Table 33, most of the verbal collocates only occur in few instances, and there is only one highly frequent collocate – *võtma* 'take', which has a log-likelihood score of 3,670.

Table 33. The verb lemmas that occur with *jalge all* (feet+under) when expressing INCEPTIVENESS

verb lemma	meaning	*n* verb in etTenTen	*n* collocate in etTenTen	log-likelihood score
võtma	take	552,338	301	3,670
olema	be	9,559,606	21	24
rulluma	roll	762	1	14
sattuma	happen (up)	60,074	1	5
seisma	stand	75,814	1	5
tulema	come	960,238	2	2
saama	get	1,871,918	1	2

In this case, the verbal collocates of *jalge all* (feet+under) express rather general meanings. Thus, it is difficult to decide on their semantic closeness. However, the extremely short list of the lemmas indicate that the use of *jalge all* (feet+under) expressing INCEPTIVENESS is quite limited. This is also indicated by the rather similar TRs. In examples (224) with the verb *võtma* 'take',

---

[64]   As shown in section 4.4 above, the total number as a complex unit is 867, *jalge all* (feet+under) behaves as a complex postposition in only 23 instances. Thus, it is not expected that the PN lemmas would suggest be productity of the complex *unit jalge* all (feet+under) in either of the functions.

(225) with *olema* 'be', and (226) with *sattuma* 'happen up', the TR is a compound ending in *tee* 'path', which all refer to a journey upon somebody. This is characteristic of most examples with *jalge all* (feet+under).

(224) *Sellegipoolest   on   seikleja-te   tee   ühine,   kuigi*
     despite   be.3SG   adventurer-PL.GEN path   common   although
     *põhjuse-d   teekonna-ø   **jalgealla**   **võtmise-ks**   ei*
     reason-PL   journey-GEN   feet.under.LAT   taking-TRL   NEG
     saa-ks-ki   erineva-ma-d   oll-a.
     could-COND-CL   different-COMP-PL   be-INF
     Lit. Despite that, the adventurers share a road, although the reasons for taking the journey under their feet couldn´t be more different.
     'Despite that, the adventurers share a road, although the reasons for taking on the journey couldn´t be more different.' [www.dragon.ee]

(225) *Muidugi   p-ol-nud   see   väga   meeldiv   sõit-a*
     of course   NEG-be-PST.PTCP   this   very   pleasant   drive-INF
     talve-ø   pimeda-s   õhtu-l   hilja   rongi-ga   kodupoole ...
     winter-GEN   dark-INE   evening-ADE   late   train-COM   homeward
     *kuid   õnneks   **ol-i**   sama   koolitee   **jalgeall***
     but   luckily   be-PST.3SG   same   school path   feet.under.LOC
     mitme-l   õpilase-l.
     many-ADE   student-ADE
     Lit. Of course it wasn't very pleasant to take the train home in the winter, dark in the night… but luckily the same road to school was under the feet of several other students as well.
     'Of course it wasn´t very pleasant to take the train home in the winter, dark in the night… but luckily there were several students who had to take on the same journey.' [naistekas.delfi.ee]

(226) *haiguse-d   hakka-vad   ne-i-d   kimbuta-ma   alles   siis*
     illness-PL   start-3PL   they-PL-PRT   perplex-SUP   only   then
     *kui   kohtutee   **jalgealla**   **satu-b**.*
     when   road to court   feet.under.LAT   happen-3SG
     Lit. They get caught up in illnesses only when they have the road to the court under their feet.
     'They get caught up in illnesses only when they are facing court.'
     [www.maaleht.ee]

The amount of productively formed examples can be observed in Figure 30, which gives the cumulative percentage of the instances *jalge all* (feet+under) expressing INCEPTIVENESS. The data show that the use of *jalge all* (feet+under) is quite restricted. The examples formed with the strongest collocate *võtma* 'take', which occurs in 301 instances, make up 91% of all examples of INCEPTIVENESS. Thus, as could also be expected given the small number of lemmas, *jalge all* (feet+under) is not productive at all as a complex unit expressing INCEPTIVENESS.
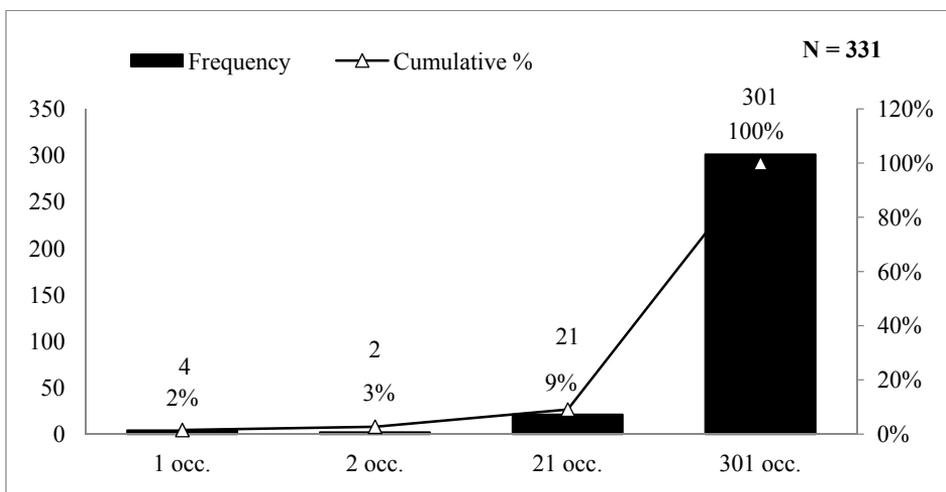
Figure 30. The cumulative percentage of examples of *jalge all* (feet+under) when expressing INCEPTIVENESS formed with verb lemmas based on their occurrence

4.6.6.2. OPPRESSION

The complex unit *jalge all* (feet+under) expresses OPPRESSION in 522 instances and takes only 8 PN lemmas and 14 verb lemmas. The very small number of PN lemmas is associated with the fact that when expressing OPPRESSION, *jalge all* (feet+under) occurs as a complex postposition only on 18 occasions. As the number of examples is extremely limited, there is no clearly distinguished group of highly frequent PN lemmas. The data suggests that the most frequent PN lemmas are *tema* 's/he' and *meie* 'we'. The rest of the lemmas occur only once or twice.

A closer analysis of the frequent collocates *tema* 's/he' and *meie* 'we' showed that in both cases we are dealing with an archaic, genre specific expressions that carry an idiomatic meaning. Such examples do not represent a productive use of the complex postpositional *jalge all* (feet+under). All of the available examples of *tema jalge all* (s/he-GEN feet+under) and *meie jalge all* (we-GEN feet+under) occur in a religious context (see examples (227) and (228)). So there is no evidence that such utterances are productively used in neutral everyday language. Rather, such usages seem to be examples of archaic expressions, such as those also found in the diachronic analysis (c.f. section 4.8.1.). Moreover, such usages seem to carry a distinct meaning. For instance, the available examples, such as (227) and (228), lack the negative connotation that is typically present in other examples where *jalge all* (feet+under) expresses OPPRESSION (see example (229)).[65] *Tema jalge all* (s/he-GEN feet

---

[65] The negative connotation comes from the perspective that is assumed when describing the inferior/superior relationship between the LM and the TR. In the case of *jalge all*

+ under) and *meie jalge all* (we-GEN feet+under) portray the same relation but in a more neutral way because the situation is construed from the perspective of the LM. Thus, *tema jalge all* (as in example 227)), was used rather to express the more specific meaning 'in one's command' and *meie jalge all* (as in example (228)) the meaning 'defeated by us'.

(227)   *Jumal*   *alista-s*   **_tema-ø_**   **_jalg-e_**   **_alla_**   *kõik*   *ning*
     God   subdue-PST.3SG   s/he-GEN   foot-PL.GEN   under.LAT   all   and
     pan-i   *tema-ø*   *kõig-i*   *asja-de*   *üle*   *pea-ks*
     put-PST.3SG   s/he-GEN   all-PL.GEN   thing-PL.GEN   over   head-TRL
     koguduse-le,   *kes*   *on*   *tema-ø*   *ihu,*   *tema-ø*
     congrecation-ALL   who   be.3SG   s/he-GEN   body   s/he-GEN
     täius,   *kes*   *täida-b*   *kõik*   *kõige-s.*
     fullness   who   complete-3SG   all   all-INE
     Lit. God subdued everything to under his feet and made him/her the head of the congregation, who is his body, his perfection, who completes him in everything.
     'God subdued everything to him and made him the head of the congregation, who is his body, his perfection, who completes him in everything.'
     [www.advent.ee]

(228)   *Ta*   *on*   *tema-st*   *vägeva-m*   *ning*   *varsti*   *talla-b*
     s/he   be.3SG   s/he-ELA   mighty-COM   and   soon   trample-3SG
     Ta   *Saatana-ø*   **_meie-ø_**   **_jalg-e_**   **_alla_**.
     s/he   Devil-GEN   we-GEN   foot-PL.GEN   under.LAT
     Lit. S/He is mightier than him and soon enough He will trample the Devil to under our feet.
     'He is mightier than him and soon enough He will trample the Devil.'
     [www.advent.ee]

(229)   *Sunnitöölis-te*   *ja*   *väljasaadetu-te*   *hulgas*   *ol-i*
     forced labourer-PL.GEN   and   deportee-PL.GEN   among   be-PST.3SG
     vähe   *vaba-sid",*   *kes*   *leid-si-d*   *ka*   *vangla-s*   *ulualus-t,*
     few   free-PL.PRT   who   find-PST-3PL   also   prison-INE   shelter-PRT
     et   *mitte*   *sattu-da*   **_okupanti-de_**   **_jalg-e_**   **_alla_**.
     that   not   happen-INF   occupier-PL.GEN   foot-PL.GEN   under.LAT
     Lit. Among the forced labourers and deportees there were very few "free" who managed to find shelter in the prison, in order not to find themselves under the feet of the occupiers
     'Among the forced labourers and deportees there were very few "free" who managed to find shelter in the prison, in order not to find themselves oppressed by occupiers.' [www.poogen.ee]

(feet+under), the situation is typically construed from the point of view of the TR (i.e. the opressee). Thus, the contextual elements (e.g. the verb, the PN lemma) are used to profile the situation negatively. For instance, in example (229), the verb *sattuma* 'happen up' is rather neutral but the PN lemma *okupant* 'occupier' carries a negative connotation.

In addition to semantic differences, there are other features that distinguish these usages. Namely, in case of *tema jalge all* (lit. 'under his/her feet') the sentential contexts are rather fixed – in the available examples *tema jalge all* co-occurs with only two verbs – *alistama* 'subdue' and *panema* 'put'. *Meie jalge all* (lit. under our feet), however, takes different verbs, such as *heitma* 'cast' (as in example 230)). Nevertheless, in both cases, the agent of the sentence always refers to God (as in examples (227), (228), and (230)). In other examples of contemporary Estonian (complex postpositions as well as complex adverbs), the agent is expressed by the LM, i.e. the 'oppressor' itself (as in 229) and (231)). Thus, *tema jalge all* and *meie jalge all* exhibits several characteristics of a fixed expression, and such examples are definitely not part of the productive use of the phrase.

(230) *Tema*   *heida-b*   *rahva-d*   *meie-ø*   *alla*   *ja*   *rahvahõimu-d*
       s/he   throw-3SG   nation-PL   we-GEN   under.LAT   and   tribe-PL
       meie-ø   ***jalg-e***   ***alla***.
       meie-GEN   foot-PL.GEN   under.LAT
       Lit. S/he throws the nations under us and the tribes under our feet.
       'S/he throws the nations under us and the tribes for us to trample on.'
       [uudised.err.ee]

(231) *Seejuures*   *kaits-ta-kse*   *näiliselt*   *küll*   *lapse-ø,*
       thereat   protect-IMPS-PRS   apparently   indeed   child-GEN
       heaolu-ø   *kuid*   *tegelikkuse-s*   *talla-ta-kse*   *las-te*
       wellbeing-PRT   but   reality-INE   trample-IMPS-PRS   child-PL.GEN
       heaolu   *võitlustandri-l*   *heitle-va-te*   ***pool-te***
       wellbeing   battlefield-ADE   struggle-PTCP-PL.GEN   side-PL.GEN
       jalg-e   ***alla***.
       foot-PL.GEN   under.LAT
       Lit. Thereat it is as if a child's wellbeing is protected but in reality the children's wellbeing is trampled under the feet of the sides fighting on the battleground.
       'Thereat it is as if a child's wellbeing is protected but in reality the children's wellbeing is trampled by the sides fighting on the battleground.' [amor.ee]

As *tema jalge all* and *meie jalge all* cannot be considered to represent productive uses of the complex postposition *jalge all* (feet+under), it is useful to determine the amount of idiomatic uses and productive uses of the complex postpositional *jalge all* (feet+under). Given that such usages make up 61% of the very few examples, the analysis based on PN lemmas suggests that *jalge all* (feet+under) is the least productive among the studied phrases.

I now turn to the analysis of verb lemmas that co-occur with *jalge all* (feet+under) expressing OPPRESSION. Similar to other phrases, most of the verb lemmas (9 out of 14) occur only once, and there are only few highly frequent lemmas. All of the verb lemmas that co-occur with *jalge all* (feet+under) in this pattern are listed in Table 34, which gives the absolute frequencies and log-likelihood scores of the verbal collocates of *jalge all* (feet+under).

Table 34. The strongest verbal collocates of *jalge all* (feet+under) when expressing OPPRESSION

verb lemma	meaning	*n* lemma in etTenTen	*n* collocate	log-likelihood score
tallama/talluma	tread	2,405	410	9,337
trampima	trample	1,405	83	1,714
tampima	pound	1,447	12	201
alistama	subdue	8,127	4	44
panama	put	347,171	4	14
sõtkuma	knead	782	1	13
materdama	bash	1,123	1	12
rõhuma	depress	3,456	1	10
olema	be	9,559,606	2	9
purustama	demolish	7,239	1	9
heitma	cast	21,152	1	6
suruma	weigh down	21,217	1	6
sattuma	happen	60,074	1	4

The most common verb is *tallama/talluma* 'tread' (as in (232)). With the log-likelihood score of 9,337, it is also without doubt a very strong collocate of *jalge all* (feet+under). The second most frequent verb *trampima* 'trample' (233) also has a strong association (1,714) with the phrase. The rest of the verbs which are less frequent also have considerably lower log-likelihood score, and are therefore much more weakly associated with the phrase. However, it is clear that most of the verbs in Table 34 are semantically very close the strongest collocates, for all of them express some sort of subduing (as in (234)–(235)). Given the semantic closeness of the verbs as well as the short list of different lemmas, it seems that the use of this pattern is rather limited and is not to be considered be quite equal with the complex units discussed above.

(232) *Kuid   praeguse-l   hetke-l   **tallu-ta-kse**   inimes-te*
    but   present-ADE   moment-ADE   tread-IMPS-PRS   people-PL.GEN
    *õiguse-d   **jalg-e**   **alla**   ja   mõne-d   piiratud*
    right-PL   foot-PL.GEN   under.LAT   and   some-PL   narrow
    maailmanägemise-ga   inimese-d   kooguta-vad   selle-le   veel
    world view-COM   people-PL   stoop-3PL   this-ALL   also
    takka.
    behind
    Lit. But at the moment people's rights are being trampled under feet and some people with a limited view of the world stoop to that.
    'But at the moment people's rights are being trampled and some people with a limited view of the world stoop to that.' [www.epl.ee]

(233) *Patoloogiline   oskus   kõik   politilise-ks   keera-ta,   kui*
pathological   skill   everything   political-TRL   turn-INF   if
kord-ki   rahvas   rohkem   oma-ø   elukorralduse-ø   ja
once-CL   people   more   own-GEN   way of life-GEN   and
*õigus-te   **jalg-e**   **alla**   **trampimise-ø**   pärast*
right-PL.GEN   foot-PL.GEN   under.LAT   trampling-GEN   because
pea-d   tõsta-b.
head-PRT   rise-3SG
Lit. The pathological skill to turn everything into a political issue, when for once
people are paying attention to their way of life and their rights being trampled
under the feet.
'The pathological skill to turn everything into a political issue, when for once
people are paying attention to their way of life and their rights being trampled.'
[www.epl.ee]

(234) *Ega   selle-st   mõtlemise-st   enam   kasu-ø   ole-ks   küll,*
nor   this-ELA   thinking-ELA   more   use-PRT   be-COND   indeed
*kui   teis-te   **jalg-e**   **all**   juba   **ole-d**.*
when   other-PL.GEN   foot-PL.GEN   under.LOC   already   be-2SG
Lit. Thinking wouldn't really help once you are under the feet of others.
'Thinking wouldn't really help once you have been trampled by others.'
[www.delfi.ee]

(235) *Ego-st   vabanemine   ei   tähenda-ø,   et   sa   luba-d*
ego-ELA   liberation   NEG   mean-CONNEG   that   you   allow-2SG
*ennas-t   teis-te-l   **jalg-e**   **alla**   **suru-da** –*
own-PRT   other-PL-ADE   foot-PL.GEN   under.LAT   weigh down-INF
ei,   sa   tea-d,   kes   sa   ole-d   ning   julge-d   ela-da
no   you   know-2SG   who   you   be-2SG   and   dare-2SG   live-INF
vastavalt   selle-le
according   this-ALL
Lit. Being liberated from your ego does not mean that you let others push you
under their feet – no, you know who you are and you have the courage to live
accordingly.
'Being liberated from your ego does not mean that you let others step on you –
no, you know who you are and you have the courage to live accordingly.'
[sisekosmos.ee]

The very high absolute frequencies of the strongest collocates suggest that the
use of *jalge all* (feet+under) as a complex unit is quite restricted. The amount of
productively formed examples can be observed in Figure 31, which gives the
cumulative percentage of the instances *jalge all* (feet+under) expressing
OPPRESSION based on the frequency of the verbal collocates.
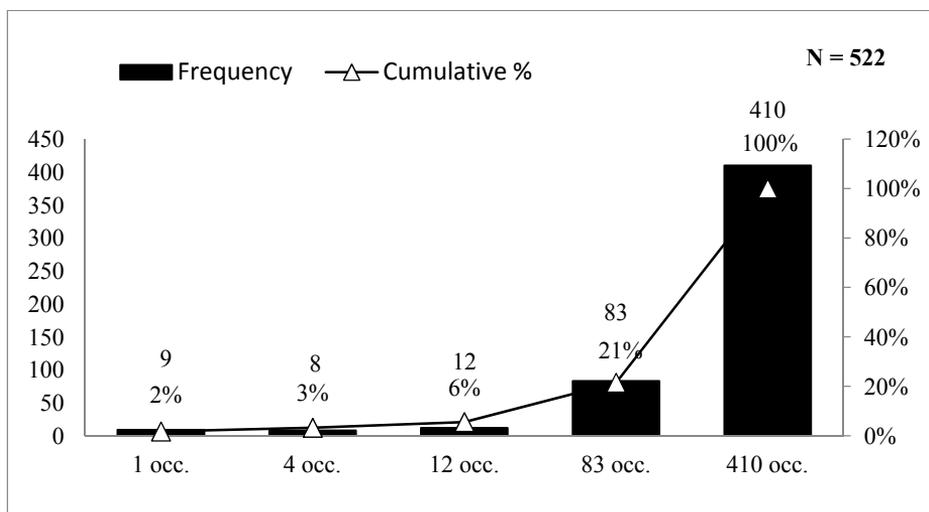
Figure 31. The cumulative percentage of examples of *jalge all* (feet+under) when expressing OPPRESSION formed with verb lemmas based on the number of occurrence

Figure 31 shows that in the case of *jalge all* (feet+under) the two strongest collocates – the highly frequent *tallama* 'tread' and less frequent but strongly associated *trampima* 'trample' – make up most of (94%) of the examples. So in this function *jalge all* (feet+under) is hardly ever used with other verbs. Thus, its usage is quite restricted and it is not considered to be productive as complex unit.

4.6.6.3. Summary of the productivity of *jalge all* (feet+under) as a complex unit

As j*alge all* (feet+under) hardly ever occurs in contexts where it is preceded by a PN, it is not particularly productive as a complex postposition. The analysis based on the the verb lemma, which is based on a larger set of data, suggests rather restricted usage in both functions – INCEPTIVENESS as well as OPPRESSION. Therefore, it can be stated that currently, *jalge all* (feet+under) resembles (a part of) an idiomatic expression rather than a function word.

4.6.7. Summary of the productivity of complex postpositions

In the previous sections, the use of the PNs and verbs with the studied phrases was examined. The data suggests that although in the case of all the studied phrases the low-frequency lemmas made up most of the usages, each phrase has also a small group of strong collocates that either form idiomatic expressions with the phrase, or more commonly are just more typical examples of the use of the studied phrases as a complex postposition. Either way, the strong collocates

are considered to form a fixed expression with the phrase, and as such cannot be considered to exemplify its productive use.

When we view the amount of fixed expressions in all of the examples where the phrases were used as complex units, it can be observed that the phrases that are more frequent are also more productive. Thus, the phrases *käe all* (hand+under) and *külje all* (side+under) are more productive – most examples (around $^{2}/_{3}$ and ¾) of *käe all* (hand+under) and *külje all* (side+under) were formed productively. The less frequent phrases *kaela peal* (neck+on) and inter-mediately frequent *jalge all* (feet+under) were less productive as the amount of productively combined examples remained below half of the examples. The use of *jalge all* (feet+under) was particularly restricted, giving evidence of its status as the least grammaticalized complex unit under study. This result is in accordance with that the other aspects of the use of *jalge all*. It is barely used as a complex postposition and is more appropriately analyzed as a part of idiomatic expressions *jalge alla tallama* 'to tread upon' and *jalge alla võtma* 'to embark upon'.

In case of the four functions of *selja taga* (back+behind) – SPACE/TIME, COVERTNESS, SUPPORT, and CONCEALMENT – and *käe kõrval* (hand+beside), there is a discrepancy between the results of PN and verb lemmas. In some cases (*selja taga* (back+behind) in functions CONCEALMENT, SUPPORT, and SPACE-TIME), the analysis of PN lemmas suggest greater productivity but verb lemmas lower productivity, in some cases (*käe kõrval* (hand+beside), *selja taga* (back+behind) expressing COVERTNESS), the other way around. Whereas somewhat lesser productivity was to be expected in less frequent phrases/functions, such as *käe kõrval* (hand+beside) and *selja* taga (back+behind) expressing SUPPORT, it is rather unexpected result in case of the spatio-temporal *selja taga* (back+behind), which is as frequent as the above men-tioned *käe all* (hand+under) and *külje all* (side+under). This result suggests that the spatio-temporal function is (still) elicited by the verb semantics.

In some cases, the most frequent verbs belong to a certain semantic classes corresponding to the semantic classes that were found among the PN lemmas of the studied phrases. For instance, the complex postpositional *käe all* (hand+under) typically co-occurs with complements that refer to authority figures (e.g. *õpetaja* 'teacher', *treener* 'coach', *juhendaja* 'supervisor') and accordingly, is used with verbs that refer to activities that can be done under the supervision of such individuals (*õppima* 'study', *treenima* 'train'). Another example is *selja taga* (back+behind) when used to express SUPPORT, which usually occurs in political discourse and, thus, co-occurs with complements referring to politicians and abstract verbs that express supporting. These examples suggest that although the studied phrases are rather productive based on their use with PN and verb lemmas, they are, in some cases still thematically restricted. Given that the development of complex function words is still in its early stages, this result is quite expected and in line with earlier research (e.g. Lehmann 1991: 503; Mikone 2000).

4.7. Summary and discussion of the results of the synchronic analysis

4.7.1. Summary of the results of synchronic analysis

Thus far, the analysis of the phrases has followed the framework outlined in the theory section, presenting the analysis criterion by criterion. In order to provide a better overview of the status of each phrase, in this section, the results of the synchronic analysis taken together and presented phrase by phrase. This is followed by a discussion (in section 4.7.2) of the results in light of previous insights into the aspects of lexicalization and grammaticalization.

KÄE ALL (hand+under). With 4401 occurrences in the etTenTen corpus, the phrase *käe all* (hand+under) is after *selja taga* (back+behind) the second most frequent of the studied phrases. The mutual information (MI) score (4.3) suggests that *käe all* (hand+under) is a strongly associated unit, the occurrence of the phrase is 16 times more likely than chance. As a complex unit, *käe all* (hand+under) expresses CONTROL, which can be further divided into physical and mental control. Similar semantic shifts have been reported in Estonian (Habicht 2000; Ojutkangas 2001) and its related languages (Ojutkangas 2001; van Pareren 2013) as well as other, non-related languages (Heine, Kuteva 2002).

Based on the proportion as well as the absolute number of the examples where the phrase is realized as the complex structure, *käe all* (hand+under) can be considered the most grammaticalized of the studied phrases. Within the 4401 available examples, *käe all* (hand+under) is realized as a complex unit on 4054 instances (92%). Moreover, as a complex unit, *käe all* (hand+under) is mostly (in 4027 instances) realized as a complex postposition and occurs as a complex adverb only on 27 occasions. *Käe all* (hand+under) occurs as a hybrid form on 79 occasions. Even though it makes up only a fraction (2%) of the data, *käe all* (hand+under) is clearly the most commonly used as a hybrid among the studied phrases.

Although *käe all* (hand+under), like the other studied phrases, is mostly used in the context of human beings, it has extended beyond human reference. Nevertheless, such examples are rare. The examples where the complex post-positional *käe all* (hand+under) is preceded by a (pro)nominal (PN) that refers to a collective or institution make up just over 1% of the data (58 instances out of 4027). Decategorialization, which is operationalized as non-agreement between the body part term and the PN was observed on 875 occasions, which make up 22% of the usages as the complex structure.

The close analysis of the verb and PN lemmas with which *käe all* (hand+under) co-occurs show that *käe all* (hand+under) as a complex unit is quite productive. Although there are certain verbs (*õppima* 'study' and *valmistama*, etc) and PN lemmas (*juhendaja* 'supervisor', *treener* 'caoach', *õpetaja* 'teacher', etc) that often collocate with *käe all* (hand+under), it occurs

mostly – 79% and 71% in case on verbs and PNs respectively – in one-off combinations or with weak collocates. However, it can be observed that as a complex unit, *käe all* (hand+under) often co-occurs with PNs that refer to authority figures. This is rather expected given that the complex item expresses CONTROL. Thus, the typical instance of the complex unit *käe all* (hand+under) is a postposition that expresses the control of an authority figure over some-body. The fact that the complex postpositional *käe all* (hand+under) is not restricted to only certain fixed usages (strong collocates) indicates high productivity of the grammaticalizing item and it is, therefore, taken to be evidence of the grammatical status of *käe all* (hand+under).

KÜLJE ALL **(side+under).** *Külje all* (side+under) occurred in the etTenTen corpus in the total of 2957 instances making it the third most frequent of the studied phrases. There is a strong association between its components – the MI score of 6.1 suggests that the phrase is a tightly bound unit. This is in line with the semantic analysis of the phrase, which suggests that *külje all* (side+under) mostly functions as a complex unit. The phrase was analyzed as a freely combined phrase on 419 occasions and a complex unit on 2530 occasions (85%). The hybrid forms make up a marginal proportion of the data (< 1%). Similar to *käe all* (hand+under), *külje all* (side+under) has a tendency to be realized as a complex postposition (2264 instances) and is less commonly (266 instances) used as an adverb. As a complex unit, *külje all* (side+under) expresses PROXIMITY, which can be further divided as physical and mental proximity. The latter is less prominent making up just 5% of all the usages of *külje all* (side+under). Similar semantic shifts have been documented in numerous other languages (see Heine, Kuteva 2002) as well as in the case of the simple function word *küljes* ('attached', 'on'; also: 'in close proximity') in Estonian.

Külje all (side+under) co-occurs with PNs of various semantic classes. In addition to animate and collective PNs that are also represented in other studied phrases, *külje all* (side+under) also co-occurs with PNs that refer to artificial and natural objects as well as to regions. However, as *külg* 'side' is also productively used as an object part (including artificial and natural objects), it is not exactly the case that the use with the wider variety of PNs is indicative of extension of the complex postposition in all cases. Rather, it seems that there are two parallel paths of extension – the usages of *külje all* (side+under) with (natural) objects have given rise to those with PNs that refer to regions, and usages with human PNs have motivated the rise of collective PNs. While the PNs that refer to region are highly frequent (2076; 92%), the collective PNs are less common (52; 2%). Thus, as a complex item *külje all* (side+under) is typically a post-position that expresses the physical proximity of an entity to a certain region. Decategorization is not well attested among the usages of *külje all* (side+under). The data suggests that the complex unit *külje all* (side+under) is displays non-agreement in 3% of the cases. Instead, such usages are a bit more common among the freely combined phrases where they made up also a marginal amount of data (7%).

Based on the analysis of verb and PN lemmas, *külje all* (side+under) is also considered to be quite productive as a complex unit. *Külje all* (side+under) is used with low frequency verb lemmas in 66% of the examples and with low-frequency PN lemmas on 72% of the examples. Due to its abundant use with PNs that refer to regions, *külje all* (side+under) is often preceded by toponyms. Such usages make up 78% of its uses as a complex postposition and are, thus, the most typical pattern of the complex postpositional *külje all* (side+under). While it is true that the toponyms increase the amount of one-off combinations of PN lemmas, it was observed that the productivity is not solely dependent on such usages. Thus, the relatively high frequency and fixedness of *külje all* (side+under), its well established extension to new contexts, and ability to be used with a variety of verb and PN lemmas are taken as features characteristic of a grammatical item.

SELJA TAGA **(back+behind).** Being the most frequent of the studied phrases, *selja taga* (back+behind) occurs on 10,958 occasions. The MI score (9.2) suggest that the components of *selja taga* (back+behind) are closely bound, which suggests fixedness. The phrase functions as a complex item on 5966 occasions, which make up just above half of all the instances of *selja taga* (back+behind) in the corpus. While there are instances of hybrid forms, they were extremely rare (19 instances; < 1%). As a complex unit, *selja taga* (back+behind) expresses four functions – SPACE/TIME, COVERTNESS, CONCEALMENT, and SUPPORT. These functions do not seem to be con-secutive stages of a single grammaticalization path but rather the results of indi-vidual developments. Therefore, they are discussed separately.

SPACE/TIME. With 4179 examples, the spatio-temporal function is the most frequent function of *selja taga* (back+behind) in the data. Such usages make up 38% of all the examples of *selja taga* (back+behind) and 70% of the 5966 examples as *selja taga* as a complex unit. The semantic shift from a body part related phrase to a spatio-temporal function word is very well attested in the world's languages (Heine, Kuteva 2002) and therefore typologically general. As a spatio-temporal word, *selja taga* (back+behind) has extended beyond human reference. The collective PNs make up almost third of the exam-ples of *selja taga* (back+behind) as a complex postposition. It should be noted that such usages occur mostly among the examples where the phrase expresses the ordinal relationship, i.e. is used most commonly when ranking the teams in sports. Decategorization was observed in 26% of the cases. Generally, the spatio-temporal *selja taga* (back+behind) prefers the adverbial function. The adverbial uses make up 92% (3860) of all the examples of this function. How-ever, despite of the fact that the spatio-temporal *selja taga* (back+behind) is rarely used as a postposition, based on the analysis of the PN lemmas, it is still rather productive (two thirds of the examples are formed productively). Never-theless, the analysis of the verb lemmas suggests that the use of *selja taga* in the SPACE/TIME function is quite restricted – in 92% of the examples the complex unit co-occurs with only two highly frequent verbs. Even though all the other

factors point to the high level of grammaticalization of the spatio-temporal *selja taga* (back+behind), the restricted use with verbs hinders that interpretation.

COVERTNESS. *Selja taga* expresses COVERTNESS on 870 occasions. Similar developments have been reported in other languages (cf. Moore 2000). Even though COVERTNESS is the second most frequent function of *selja taga* (back+behind) as a complex unit, it is still far less frequent than the spatio-temporal function discussed above. When expressing COVERTNESS, *selja taga* (back+behind) is equally frequent as an adverb (54%) and as a postposition (46%). As a complex postposition expressing COVERTNESS, *selja taga* has extended beyond human reference – the collective PNs appear in 21% of the examples. Decategorization was observed in 27% of the examples. As for the productivity, the analysis of the PN lemmas and verb lemmas give different results. The analysis of PNs indicates that about half of the examples of COVERTNESS are formed productively. The analysis of verb lemmas indicates that 79% of the examples are formed productively. Thus, in this case, the re-occurring nominal complements suggest that the use of the complex post-position is (still) somewhat restricted.

SUPPORT. *Selja taga* (back+behind) occurs as a complex unit expressing SUPPORT on 610 occasions. Thus, SUPPORT is not a particularly frequent function of *selja taga* but still forms a clearly distinct pattern, which is also attested in other languages. In this function, *selja taga* (back+behind) is inclined to be realized as a complex postposition (80%). As a postposition *selja taga* (back+behind) is complemented by a collective (pro)nominal in 17% of the instances, which is the lowest level among the individual functions of *selja taga* (back+behind). The complex postposition displays decategorization in 25% on the cases. The analysis of the PN lemmas suggests that about ¾ of the examples exhibit productive patterns. However, the data also indicate that productivity of *selja taga* (back+behind) is somewhat dependent on its ample use with proper names. Moreover, the analysis of the verbal collocates suggests quite restricted use (36% of productively formed examples). Thus, in accordance with its low frequency, the productivity of the complex postposition expressing SUPPORT does not indicate a particularly high level of grammaticalization.

CONCEALMENT. With just 302 occurrences as a complex function word, CONCEALMENT is the least frequent among the functions of *selja taga* (back+behind). When expressing CONCEALMENT, *selja taga* (back+behind) is inclined to be realized as a postposition (92%). As a complex postposition expressing CONCEALMENT, *selja taga* (back+behind) exhibits decategorization on 20% of its instances and extension on 21% of its instances. It is notable that in this function, *selja taga* (back+behind) is often (23%) com-plemented by a noun that refers to God. The analysis of the individual lemmas suggests that in such usages, *selja taga* (back+behind) is part of an idiomatic expression (*nagu*) *vanajumala selja taga* 'secure, protected'. Moreover, the phrase is drawn to certain strong verbal collocates in over half of the examples, which suggest that the use of *selja taga* when expressing CONCEALMENT only semi-productive.

KÄE KÕRVAL (hand+beside). *Käe kõrval*, which is one the least frequent phrases in this study, occurs on 780 instances in etTenTen. Its MI score suggests that the components of the phrase are strongly associated and that the likelihood of the components occurring as a phrase is above chance. As a complex unit, *käe kõrval* (hand+beside) expresses functions BESIDE and ACCOMPANIMENT. As a complex unit, *käe kõrval* (hand+beside) is inclined to occur as an adverb. Complex adverbial uses make up 84% of its occurrences. Because the amount of complex postpositions is very small, there are very few examples that suggest extension (1%) and decategorization (11%, 13 examples). The analysis of the individual PN and lemmas suggest that as a complex unit, *käe kõrval* (hand+beside) is somewhat restricted – about half of the examples are formed productively. However, based on the analysis of the verbal collocates, two thirds of the complex units *käe kõrval* (hand+beside) are formed productively.

KAELA PEAL (neck+on). As the least frequent of the studied phrases, *kaela peal* (neck+on) occurs in etTenTen on just 216 occasions. However, it is a strongly associated phrase – the MI score suggests that the likelihood of its components occurring as a phrase is almost 8 times above chance. As a complex unit, it expresses the rather specific function BURDEN. Such a shift does not suggest typological generality. The phrase occurs as a complex unit on 87 occasions (41% of the examples). As a complex unit, *kaela peal* (neck+on) has a slight tendency to occur as a postposition (59% of the cases). Despite of its low frequency and specific meaning, the data suggest that both extension and decategorization are well established among the complex postpositional *kaela peal* (neck+on). Extension beyond human reference was detected in 14% and decategorization among 41% of the examples of complex postpositions. However, the analysis of the PN lemmas suggest that over half of the examples are formed with strong collocates, and the analysis of verb lemmas suggest that as much as two thirds of the examples include strong collocates. These restricted pattern do not suggest that *kaela peal* (neck+on) is a productive grammatical item.

JALGE ALL (feet+under). With 1918 occurrences in the etTenTen corpus, *jalge all* (feet+under) is intermediately frequent among the studied phrases. Regarding the collocational strength between the components of the phrase, *jalge all* (feet+under) can be considered to be highly fixed (MI = 4.7). It occurs as a complex unit on 867 occasions. As a complex unit, it expresses two functions – OPPRESSION and INCEPTIVENESS, which are considered to be results of different instances of lexicalization. Both of these functions are rather specific and do not suggest typological generality of these developments. This is indicative of lexicalization rather than grammaticalization. In both functions, *jalge all* (feet+under) clearly prefers to be realized as an adverb. The postpositional uses make up just 4% (in case of OPPRESSION) and 2% (in case of INCEPTIVENESS) of the complex units. In accordance with the tendency to be realized as a complex adverb, *jalge all* (feet+under) has not extended beyond human reference (except for in some idiomatic expressions which do not

suggest grammaticalization). The decategorization is not observable in case of *jalge all* (feet+under) because its first component is in plural. The analysis of the individual lemmas suggests that the available examples of *jalge all* (feet+under) mostly (94%) consist of fixed usages. Therefore, *jalge all* (feet+under) is currently rather a part of (semi-productive) idiomatic expressions *jalge all tallama/trampima* 'tread upon' and (*teed*) *jalge alla võtma* 'embark upon'.

In sum, based on the results of synchronic analysis, *käe all* (hand+under) and *külje all* (side+under) are considered to be the most grammaticalized of the studied phrases, and *jalge all* (feet+under) the least grammaticalized. The rest of the complex units are located between them on the lexicon-grammar continuum because the different factors analyzed point to different directions.

4.7.2. Discussion of the results of synchronic analysis

As the development of Estonian complex postpositions constitutes a development of a new contentful form, which is not directly derivable from its components, it is in line with the definition of lexicalization by Brinton and Traugott (2005: 96). However, Brinton and Traugott consider the development of complex function words, such as complex prepositions, to be an instance of grammaticalization instead (2005: 65). Drawing from Lehmann (2001; 2002), Rostila (2004) and many other students of complex adpositions, the present study considers the process at hand to be an instance of lexicalization as well as grammaticalization. The development of Estonian complex function words is considered to be an instance of grammaticalization because the output of the process is a postposition, i.e. a grammatical item. However, the development of such complex units also includes lexicalization because it involves a development a freely combined complex unit into a holistically processed unit (Lehmann 2002: 1–2).

The analysis shows that, in general, the studied phrases show little formal evidence of the actualization of reanalysis of the freely combined phrases into complex postpositions. Of course, this result was to a certain extent anticipated as the present study investigates a language change which is clearly an instance of incipient grammaticalization. Moreover, the development of complex function words, especially the development of complex postpositions, has received very little attention so far, and according to the traditional view, the category of postpositions consists of simple forms only. Therefore, it is quite expected that the studied phrases lack characteristics that would suggest a high level grammatical status. Nevertheless, in each of the studied phrases, some evidence, including formal evidence, of grammaticalization was found. All of the studied phrases (except for *jalge all* (feet+under) have extended beyond human reference. That being said, it must be reiterated that the examples that display extension are still rather infrequent – the collective (and to a certain extent, abstract) PNs are most common among the examples of *selja taga* (back+behind) and

kaela peal (neck+on), where their proportion remains around 20% of the examples. All of the phrases also displayed decategorization, which was operationalized as the non-agreement in number between the PN and the body part term. In general, non-agreement is more extensively established among the studied phrases than extension (3%–31%) but its association with the complex forms was not as evident as in case of collective PNs, and overall, the amount of such examples was still relatively small. Thus, even though both of these parameters are certainly inclined to be established in the complex structure, it is only the semantics that discriminates between the simple and the complex structure.

This type of semantic change seems to be characteristic of grammaticalization of Estonian function words in general. As there are not many formal features that separate the source and the target forms, the development into a (complex) holistic unit is primarily a semantic change. For instance, in the case of English complex prepositions, one the key indicators of the status of the complex preposition would be the loss of the article in the PNP constructions (cf. Svorou 1994; Hoffmann 2005). For instance, Hoffmann (2005: 56) suggests that the drop of the definite article in *in view of* is indicative of decategorialization of *view*. However, because Estonian does not make use of articles, this feature cannot be observed. Instead, the complex function words develop through lexicalization of the noun and the simple postposition (*käe+kõrval* ('hand'+'next to') > *käekõrval* 'beside, accompanied (by)'). This process is analogous to that of the simple function words in Estonian, which have (mostly) developed through a similar process, whereby a noun and a locative case suffix form a semantically holistic unit (*kõrva+l* ('ear'+ADE) > *kõrval* 'next to'). Therefore, as there are few formal factors to differentiate between the source and the target form, the semantic factor has more weight.

The prominent role of lexicalization in the development of complex postpositions in Estonian is further emphasized by the fact that the complex items under study mostly occur in the context of human beings. As suggested above, the inclination towards human beings, on the one hand, indicates that the development of complex units is still in its early stages. However, the grammaticalization of body part terms is usually described as a path starting with a body part term, which is then extended to refer to object part and further on to spatial grams (e.g. Svorou 1994: 90; Heine 1997: 44). For instance Svorou (1994: 90–91) describes the development of the complex preposition *in front of* in the following pathway: *front* (body part) > *the front of* (object part) > *in the front of* (location in contact with the object part) > *in the front of* (location near the object part) > *in front of* (complex preposition). The data of Estonian complex postpositions shows that in most cases, the body part term is not productively used to refer to object parts at all but the phrase behaves as a complex item nevertheless. The fact that most of the studied body part terms are not used to refer to object parts is perhaps not that striking by itself. However, what I would like to point out is that in this case, the grammaticalization is not facilitated by extension of the body part term to new contexts (first to object part and

then to spatial reference) but by the lexicalization of the body part related phrase as a whole. Once the phrase is lexicalized and it goes under desemanticization, it may be extended to new contexts as a whole.

Indeed, the data show that as complex units, the phrases have gone under desemanticization. This is suggested by the fact that as complex units, the body part terms have lost their referential capacity, and as whole, the phrases express more abstract meanings than their source form. Nevertheless, in some cases, the meanings of the complex items are rather specific. The specificity is constituted by the fact that it is difficult to find parallel shifts in grammaticalization in other languages, as was suggested for *jalge all* (feet+under), *kaela peal* (neck+on), (and to some extent, *käe kõrval* (hand+beside)) (see sections 4.3.5, 4.3.6, and 4.3.4). Thus, even though these meanings are abstract they are not considered to be highly desemanticized, but rather quite idiosyncratic, which is characteristic of lexicalization, not grammaticalization (Brinton, Traugott 2005: 97).

As suggested by Brinton and Traugott (2005: 18) the status of a lexical vs. grammatical item is also dependent on the productivity of the linguistic item. While grammaticalization is usually associated with higher productivity and lexicalization lower productivity (Brinton, Traugott 2005: 100; 97), it is interesting to observe the productivity of items that involve both of these phenomena. The data show that while frequency of a linguistic item does not equal productivity, the two notions are related – the more instances the linguistic item occurs in, the richer its sentential context. In general, it was observed that the more frequent phrases (*käe all* (hand+under) and *külje all* (side+under)) are also more productive, the less frequent (*kaela peal* (neck+on), *käe kõrval* (hand+beside)), but especially the intermediately frequent *jalge all* (feet+under) were less productive. *Käe kõrval* (hand+beside) and the various functions of *selja taga* (back+behind) can be considered as semi-productive. Thus, it seems that the more specific phrases are also less productive, which is also indicative of their lexico-grammatical status. Therefore, the less productive phrases of higher specify are considered to be lexicalized but not particularly far in the process of grammaticalization yet. Therefore, on the lexicon-grammar continuum, they are closer to lexical items and the more frequent and productive phrases are closer to grammar.

However, the distinction between lexical/grammatical and productive/unproductive items is not perceived as discreet. Brinton and Traugott point out that much like lexical and grammatical items are viewed as two extremities of the same cline, the productivity of a linguistic item best described as a continuum as well (2005: 16). For instance, the spatio-temporal *selja taga* (back+behind) is a frequent item that carries an abstract meaning and is productive based on the analysis of PNs. Yet, on more than 90% of the cases it is combined with two verb lemmas. The latter does not suggest a high level of grammaticalization even though the other factors point to this direction. This discrepancy between the factors is taken to be indicative of the fact that the spatio-temporal *selja taga* (back+behind) is lexicalized and going under desemanticization but the contextual expansion is yet to happen. When the

studied phrases are viewed in the context of the three-stage clines of grammaticalization and lexicalization proposed by Brinton and Traugott (2005: 101–102), the more frequent and productive phrases can be located between the levels G1 and G2 because they behave as function words, which are (as post-positions) syntactically dependent on their nominal complements. I would not consider them (yet) semi-bound units (G2) and definitely not affixes (G3). The less frequent and less productive and more specific phrases are analyzable as ranging between the levels L1 and L2 because they are definitely fixed complex forms with an independent meaning, yet they are still transparent. The status towards the levels L2 and G2 is indicated by the fixedness of these phrases.

The analysis showed that all of the studied body part related phrases are fossilized in the singular form. The fixedness of the Estonian complex postpositions may be further facilitated by their structure. The complex phrase consists of a body part noun in the genitive case and a simple postposition (N-SG.GEN+P). As was already stated above, there are no articles within the phrase, which makes the phrase less complex. The articles serve as markers of boundaries between the words as well as preserve the perception of the phrase as a syntactic structure. Of course, Estonian has other means to mark the boundaries of words, for example, case suffixes. Nevertheless, within the postpositional phrases, there are no demarcation elements because the genitive ending has been lost. The fact that there is no case ending to explicitly mark the borders of the two words can be seen as another factor that contributes to the holistic interpretation of postpositional phrases in Estonian. A similar proposal has been put forward by Grünthal who suggests that the adjacency and exclusion of additional morphemes between the noun and postposition underline the density within the postpositional phrase (Grünthal 2003: 109). Kabak (2006: 35) who has discussed the role of case suffixes in the similar type of change in Turkish postpositions claims that the absence of case suffixes may contribute to the morphologization because the lack of morphological marker makes the border of the noun and the postposition unclear.

Despite of the observed fixedness and of the studied phrases, in some cases there are still evidence that the phrases are analyzed as freely combined items. This is especially evident in cases where the first component of the phrase, i.e. the body part term is pre-modified by and adjective. Hoffmann (2005: 78) observed frequent adjectival modification in the data of the *in need of* (e.g. *in desperate need of*), a complex prepositions that displays no semantic shift. In the present study, however, the adjectival modifiers are mostly present in cases where the phrases express the meanings that they express as complex items. Such, so-called hybrid forms are rather infrequent in my data. They are most common among the usages of *käe all* (hand+under) where they make up only 2% of the data (79 examples). As the noun *käsi* 'hand' is often modified by an adjective yielding such expressions as *kuldsed käed* 'golden hands' or *abistav käsi* 'helping hand', it is likely that such patterns have also been carried over to usages where *käsi* 'hand' is followed by a postposition *all* (under). However, the tendency to be used with adjectives is probably not dependent on the noun

only because the data of the phrase *käe kõrval* (hand+beside) does not include more than two instances of hybrid forms. Therefore, it seems that the preceding adjectives are compatible with *käe all* (hand+under) as a whole. In fact, it is quite likely that the presence of the adjective supports the development of the complex unit because the adjective helps to project the meaning of the complex unit. Adjectives such as *range* 'strict', *hoolas* 'diligent' bring forth the sense of control that is expressed by the complex unit because the adjectives refer to features that are not so much characteristic of the hand as a limb but the human as the 'owner' of the limb and thereby also to the type of relationship (strict, diligent, etc.) the body part related expression projects. Thus, the hybrid forms can be seen as an intermediate stage between the simple and complex structure. However, because not all of the studied phrases productively occur with adjectives, the usages with hybrid forms do not seem to be a vital stage in the development of complex function words.

In general, the existence of forms that exhibit elements of the complex and the simple structure could suggest an intermediate stage between the two possible analyses of the phrase. This phenomenon has been discussed in De Smedt (2012) who suggests that the existence of hybrid forms is at odds with the abrupt nature of reanalysis. Haspelmath (2011) argues that reanalysis is not abrupt but a gradual change, which is in line with the gradient nature of morphology and syntax. Unfortunately, the Estonian data does not provide much evidence to the dispute on the abruptness of reanalysis. In this case, the hybrid forms are rather infrequent, they do not seem to be a vital stage in the grammaticalization path, and in essence, they do not necessarily suggest a contradiction between two alternative categories but a figurative use of the simple structure (which may facilitate the development of the complex structure). However, many other factors observed in this study suggest that language change does not happen abruptly and that there are several interpretations available at the same time. For instance, non-agreement was not particularly systematically associated with the complex form, the different types of productivities produced at times different results, a low-frequency postposition *kaela peal* (neck+on) was one of the two phrases most commonly extended beyond human reference while the intermediately frequent *jalge all* (feet+under) is barely a complex function word at all. As all of the factors under study do not (always) point to the same direction, it seems natural that the items cannot be analyzed as one or the other structure in each and every usage. By this, I do not mean to claim that reanalysis does not exist but rather that language users process and produce language in larger meaningful units, which do not necessary fit into one or other category at every instance of use.

4.8. Diachronic development of body part related complex function words

4.8.1. Introduction

Although studies of grammaticalization often include synchronic analysis, diachronic changes are of particular interest. Grammaticalization is a language change that occurs gradually over time, hence, the process of grammaticalization is mostly studied diachronically. Moreover, in order to show that there has been a change in language, it is necessary to compare different points in time. Thus, in this chapter, I present the results of a diachronic analysis of body part related postpositional phrases. Owing to the possible methodological problems discussed in section 3.2.2, the results of the analysis should be interpreted with caution.

The diachronic data comes from three corpora – COLE, 19th century texts, and CELL – which were introduced in section 3.2.3. In order to investigate the dynamics of the usage of each phrase, I analyze the usage of the studied phrases in three periods of time – 17th–19th century (COLE), the first half of the 20th century (the sub-corpora of the 1890s, 1900s, 1910s, 1930s and 1950s) and the second half of the 20th century (the sub-corpora of the 1960s, 1970s, 1980s and 1990s). The periodization is based on the length of the periods these corpora cover, the size of the sub-corpora, as well as the compatibility of the texts with regard to the genre (see section 3.2.2). In order to be able to compare data from different periods, the data is presented as occurrences per million words (pmw).

The aim of the diachronic study is to gain insight into the very beginning of the development of complex function words studied here and to describe the grammaticalization paths of body part related complex function words as they have evolved over time. Thus, the study addresses the following questions:

1. Is there evidence of increasing grammaticalization?

In the diachronic data, I analyze the same aspects as in the synchronic data, i.e. lexicalization, extension beyond human reference (inanimate complements of the complex postpositions), decategorialization (non-agreement between the preceding noun and the body part term), and text frequency and function frequency of the studied phrases (see sections 2.5.1–2.5.4). However, the diachronic data also allows me to observe any changes in the frequency of the studied phrases over time. Thus, in the diachronic analysis, I observe the overall frequency of the studied phrases and the frequency of each of the parameters listed above. In line with the general principles of grammaticalization, I expect an increase in overall frequency[66] as well as in lexicalized usages and instances

---

[66] However, according to Mair (2004) an increase in overall frequency should not be expected in the case of incipient grammaticalization. Hoffmann (2005) proposes that high frequency is not necessary for grammaticalization to occur (see section 2.5.4).

that manifest extension and decategorialization. However, because the corpora are quite small and the data few, it is not possible to meaningfully observe any increase of productivity in the studied phrases. Thus, the lemmas of verbs and PNs are not observed in the diachronic data.

2. What was the diachronic order of changes?

Habicht and Penjam (2007: 57) have proposed that Estonian function words that serve as complex adverbs as well as complex postpositions followed the developmental path:

Noun form > simple adverb > simple postposition > complex adverb > complex postposition

Their suggestion is based on observations that have been made about simple function words (Habicht 2000: 50). Habicht and Penjam (2007: 56) hypothesize that complex function words follow the same path. The synchronic analysis allowed me to observe the adverbial and postpositional uses based only on their frequencies in the studied phrases (see section 4.4). The diachronic data however, allows me to observe the diachronic order of appearance of the adverbial and postpositional usages. Thus, in this study, I attempt to find diachronic evidence that support or refute the hypothesis.

3. What motivates such changes?

Another point of interest is the nature of the grammaticalization of body part related postpositional phrases. Grammaticalization may constitute a language-internal change or a contact-induced change (cf. Heine, Kuteva 2005). It has been claimed that the development of certain Estonian simple function words that entered the grammaticalization process in the 16th–17th century has been motivated by the influence of the German language (Habicht 2000: 51–52). The possibility of German influence has not been studied so far regarding the development of complex function words. There is no set list of criteria to determine foreign influence in the process of grammaticalization. Thus, I have used the parallel texts in German that are provided for some of the texts in COLE, to determine whether the Estonian phrases are used as replicas of similar complex grammatical(izing) functions in German.

Diachronic analysis for each of the studied phrases is presented one by one. Where appropriate, the results of the diachronic analysis are contrasted with that of the synchronic analysis. However, it should be noted that due to the vast differences in the size of the corpora (see sections 3.2.1.1 and 3.2.3), the numeric results of the diachronic and synchronic analyses are not directly comparable. The presentation of the analysis of each of the phrases is structured as follows. First, I present the distribution of the free and complex units, as well as the

distribution of the individual functions, in each of the observed periods; second, I discuss the distribution of the adverbial and postpositional usages in the observed periods, and address the question of the diachronic order of the adverbial and postpositional stages in the developmental path of the complex function words; third, I discuss the contextual expansion of the studied phrases, which is operationalized as the properties of semantic class and grammatical number of the PN; last, I present a summary of the results.

4.8.2. *Käe all* (hand+under)

The first attestations of *käe all* (hand+under) in the diachronic data date back to the year 1603, when the phrase was used by Georg Müller in his sermons (see example (236)). It can already be observed from these first attestations that *käe all* (hand+under) bears non-literal meanings. For instance, in (236) *käe all* (hand+under) is used to express meanings that are close to the notion of MENTAL CONTROL that the phrase carries in contemporary language (see section 4.3.1).

(236) *Sÿß   tæña-Ø   sinu-Ø   Su-Ø   ninck Süddame-Ø   kaas,*
then   thank-IMP   you-GEN   mouth-GEN   and   heart-GEN   with
ninck   lasc-ke-m   meÿe   sÿß   hen-d   igka   aÿa-l
and   let-IMP-1PL   we   then   ourselves-PRT   every   time-ADE
*Iumala-Ø   tæma-Ø   keicke-ø   wægkewa-Ø   **Kæe-Ø   alla***
God-GEN   s/he-GEN   all-GEN   *mighty-GEN*   hand-GEN   under.LAT
allanda-da   sÿß   pidda-b   sesame   keick   mea   meÿe   the-me,
humble-INF   then   must-3SG   same   all   what   we   do-1PL
Iumala-Ø   mele-s   hæ   ninck   armas   olle-ma.
God-GEN   mind-INE   good   and   lovely   be-SUP
'Then thank with your mouth and heart and let us humble ourselves under the hand of God all the time, then everything we do will be good and lovely in the eyes of God.' [COLE, Müller, 1603, s. 9, 11]

Unfortunately, there is not much diachronic data available. The total number of examples of *käe all* (hand+under) over the four centuries is 96. Due to the lack of data, it should be noted that the following analysis should be interpreted with caution. However, the available data helps me to reveal the developmental path in many respects. The findings from the diachronic corpora are discussed qualitatively and, as much as the data allows, quantitatively. The analysis will follow the structure introduced in section 4.8.1 above.

In the following, I discuss the overall frequency of *käe all* (hand+under) in the diachronic data and the distribution of the simple and complex structures, as well as the hybrid forms, in the observed periods. The overall frequency is presented in Figure 33.
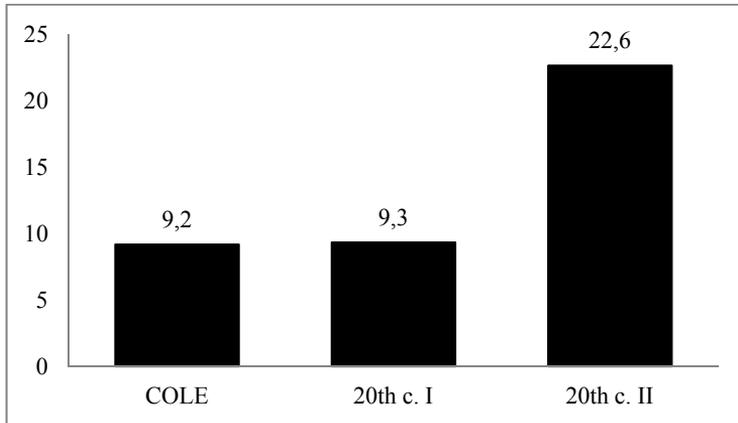
Figure 33. The frequency (pmw) of *käe all* (hand+under) in the observed time periods

The data show an overall frequency increase of *käe all* (hand+under) over the observed periods. In COLE and the 1st half of the 20th century, the frequency remains around 9 instances pmw, but by the 2nd half of the 20th century, it has increased to 22.6 instances pmw. As an increase in the frequency of an utterance is generally considered to co-occur with the process of grammaticalization, the results for *käe all* (hand+under) are consistent with the general principles of grammaticalization. However, Mair (2004: 138) suggests that the frequency of grammaticalizing items should be viewed in relevant contexts. Thus, in the following, I present the frequency of *käe all* (hand+under) as a simple structure, a complex structure, and as hybrid forms. When observing the frequencies of the individual structures, two aspects are of particular interest – first, the first attestations of the structures and second, the dynamics of the frequencies of the structures during the observed periods. Additionally, I will touch upon the distribution of the structures within each period. The frequencies of the structures are presented in Figure 34.
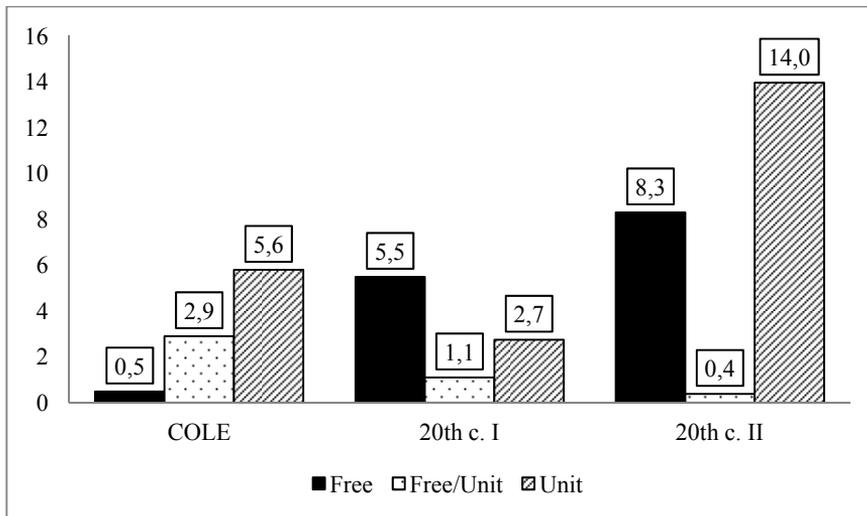
Figure 34. The distribution of the simple and complex structure as well as hybrid forms of *käe all* (hand+under) in the observed periods as instances pmw

The data show that all of the structures were already present in COLE, i.e. the freely combined phrases, complex units, and hybrid forms were attested in Old Literary Estonian. Figure 34 shows that along with an increase in the general frequency of the phrase, the relative frequency of the freely combined phrases increases throughout the observed periods. In COLE, there are 0.5 instances (pmw) of freely combined phrases, which has risen up to 5.5 instances pmw in the 1st half of the 20th century and up to 8.3 in the 2nd half. The frequency of the hybrid forms however, is decreasing over the observed periods. COLE includes 2.9 instances of hybrid forms, but there are 1.1 instances (pmw) and 0.4 instances (pmw) in the data for the 1st half of the 20th century. The frequency of the complex units does not seem to have a clear direction. In COLE the complex units are rather frequent, amounting to 5.5 instances pmw. However, in the data for the 1st half of the 20th century, the frequency of the complex units decreases to 2.7 instances (pmw). In the 2nd half of the 20th century, the frequency of the complex structures increases considerably to 14 instances pmw. In the following these results will be interpreted, starting with the analysis of hybrid forms.

As was suggested in section 4.2, hybrid forms exhibit semantic features similar to that of complex units and structural features similar to freely combined phrases. For instance, in (237) *käe all* (hand+under) bears a similar meaning to that of contemporary MENTAL CONTROL. However, the adjectival modifier (*wegkiwa* 'mighty') precludes its analysis as a complex postposition. Semantically, the adjective reinforces the non-literal interpretation of the phrase. Thus, diachronically, the hybrid forms may be considered as an intermediate stage between the simple and the complex structure. Therefore, it is quite natural that the frequency of such usages decreases as *käe all* (hand+under) gives in to complex postpositional interpretation. Moreover, all of

the examples of hybrid forms in the data (with one exception in the 2nd half of the 20th century) are of very similar usages (e.g. as in (236) and (237)). Indeed, a closer analysis on these examples reveals that they are extremely restricted. In addition to the adjectival modifier of the body part term, such usages of *käe all* (hand+under) can only be found in the lative form (*käe alla*); they were used with only one verb (*end*) *alandama* 'to humble (oneself)'; the (pro)noun preceding the phrase only refers to God. As such, examples like (236) and (237) could be analyzed as instances of an idiomatic expression.

(237) *SJhs*   *allanda-ke-t*   *hen-d*   *nühdt*   ***Jummala-ø***
      then   humble-IMP.2PL   ourselves-PRT   now   God-GEN
      wegkiwa-ø   ***kehje-ø***   ***alla***   *eth*   *temma*   *te-i-d*
      mighty-GEN   hand-GEN   under.LAT   so   s/he   you-PL-PRT
      üllenda-p.   *omma-l*   *aja-l*
      promote-3SG   own-ADE   time-ADE
      'Then humble yourselves under God's mighty hand, that He may lift you up in due time.' [COLE, Stahl, 1638, 96]

Moreover, there is evidence that such usages might have developed due to influence of the German language. Considering example (238), the parallel text in German (239) contains the same exact expression (*unter seine gewaltige Hand* 'under his mighty hand'). As the authors of these examples are both German, it is quite possible that examples in Estonian were direct translations from German. In addition, the translations of religious texts in general are rather exact.

(238) *Meije*   *pea-me*   *Jummala-Ø*   *deni-ma*   *Allandickult*   *eth*
      we   must-1PL   God-GEN   serve-SUP   humbly   so
      meije   *hen-d*   ***temma-Ø***   ***wegkiwa-Ø***   ***kehje-Ø***   *alla*
      we   ourselves-PRT   s/he-GEN   mighty-GEN   hand-GEN   under.LAT
      allanda-me.
      humble-1PL.
      'We must be humble servants of God. Therefore, humble yourselves under the mighty hand of God.' [COLE, Stahl, 1641, 110]

(239) *Wir*   *sollen*   *GOTT*   *dienen*   *Demütiglich*   *das*   *wir*   *vns*
      we   *must*   *God*   *serve*   *humbly*   *so*   *we*   *ouselves*
      vnter   ***seine***   ***gewaltige***   ***Hand***   ***demütigen***
      under   his   mighty   hand   humble
      'We must be humble servants of God. Therefore, humble yourselves under the mighty hand of God.' [COLE, Stahl, 1641, 110]

Among the very first instances of *käe all* (hand+under) in the diachronic data, there are also examples that occur without a preceding adjective. Despite the fact that such instances are close to the examples discussed above (in examples (237) and (238)), they are, based on their semantics and phrasal structure, analyzed as complex units. For instance in example (240), which originates from 1605, *käe all*

(hand+under) is used with the same PN (*Jumala* 'God') and verb (*alandama* 'humble') as in the examples of hybrid forms. However, they are also close to the complex postpositional usages that express MENTAL CONTROL in contemporary language. Moreover, in the 18th century, different contexts also emerge, as in example (241) where PN refers to a priest's son and the verb is *olema* 'be'. Nonetheless the general context of the use of *käe all* (hand+under) remains close to that of the hybrid forms. Based on the available data, it seems possible that the postpositional uses have developed from the 17th century hybrid forms. Mikone (2000) also suggests that grammatical items may rise from idiomatic expressions. Analysizing the relationship between idioms and Estonian postpositions, she claims that due to frequent use, idioms lose their stylistic markedness, and may be extended to further contexts and, consequently, become less fixed. The 'backbone' of the idiom remains fixed but the rest of the expression may vary given that the variation follwos certain semantic requirements. (Mikone 2000: 25). In this case, the 'backbone' would be *käe all* (hand+under), which in time may have taken on different PNs and verbs as exemplified in (241).

   Moreover, if the postposition *käe all* (hand+under) has risen due to German influence, it would follow that the beginnings of the complex postpositional *käe all* (hand+under) could be considered as contact induced grammaticalization. However, it should be noted that in the case of *käe all* (hand+under) it does not seem to be an instance of grammatical copying in the sense of Heine and Kuteva (2005), because the source *unter deine gewaltige Hand* ('under his mighty hand') is not a grammatical item (e.g. a complex gram), but a figure of speech. Thus, it is not the grammatical construction that is borrowed, but only the meaning of an idiomatic expression.

(240)   *... Kuÿ meÿe hen-d   keicke-st Süddame-st **Iumala-ø   Kæe-ø***
         if   we   ourselves-PRT   all-ELA   heart-ELA   God-GEN   hand-GEN
         alla   *{hen-d}*   *allanda-me, sÿß   on   mödewarsÿ*
         under.LAT   {ourselves-PRT}   humble-1PL then be.3SG   shortly
         Iumala-ø   nuchtlus   kebie-mb   v¨lle   me-i-dt.
         God-GEN   judgement easy-COMP   over   we-PL-PRT
         'If we humble ourselves under the hand of God with all our hearts then shortly God's judgement over us will be easier.' [COLE, Müller, 1605 s. 28, 4]

(241)   *Se   on   Kersoni-ø   poega-de   sugguwössade   tenistus*
         this be.3SG   Kerson-GEN   son-PL.GEN   family tree-PL.GEN   service
         koggodusse-ø   telgi-ø   kallal, ja   mis   ne-i-le
         congregation-GEN tent-GEN   at   and   what they-PL-ALL
         *tulle-b   ärratehj-a,   pea-b   **olle-ma   Jtamari-ø   preestri-ø***
         must-3SG do-INF   must-3SG   be-SUP   Jtamar-GEN   priest-GEN
         Aaroni-ø   poia-ø   käe-ø   al.
         Aaron-GEN   son-GEN   hand-GEN   under.LOC
         'This ist he service of the sons of Kerson on the congregation tent and what they must do must be under the hand of Aaron, the priest of Jtamar.'
         [COLE, 1739, 4 Mo 4:28]

The contextual attachment to religious discourse may also serve as an explanation of the fluctuating dynamics of the *käe all* (hand+under) as a complex unit. That is, the decrease in frequency of the complex items in the 1st half of the 20th century may be due to a genre difference between the corpora. As discussed in sections 3.2.2 and 3.2.3, COLE is mostly compiled of religious texts, which are not represented as an individual genre in the source (CELL) of the 20th century texts. Thus, it may be the case that *käe all* (hand+under) was contextually over represented in COLE and that the 20th century data reflects its development in more neutral discourse.

Last, I discuss the frequency of the freely combined phrases in the observed periods. According to the general principles of grammaticalization, one would expect to see a directed change over time in the distribution of the usage functions from freely combined units to fixed units, i.e. an increase in the frequency of complex units and a decrease in frequency of freely combined phrases. However, the data suggests that the simple structure increases over the centuries. This is not taken to suggest that *käe all* (hand+under) does not adhere to the general principles of language change, but is rather considered to be a concurrence of several factors. First, it is likely that the low frequency of the literal uses of the phrase in COLE is connected with the genre differences between the two corpora. It is natural that there are few literal usage of *käe all* (hand+under) in COLE, which mostly consists of religious texts rich in figurative language, and therefore literal uses might not have shown up in the corpus. In addition, the high specificity of the literal interpretation of *käe all* (hand+under) might be at fault as well. The increase of the simple structure in the observed periods can be explained by the fact that the 20th century data also includes instances of *käe all* (hand+under) that are part of a fixed expression *käe alt kinni hoidma* (lit. to hold beneath one's arm '[to be] arm-in-arm') but nevertheless, express literal meanings (as in (242)).

(242) *Tüdruk,*   *kelle-Ø*   *nägu*   *jä-i*   *pimeduse-ø*   *varju,*
       girl   who-GEN   face   remain-PST.3SG   darkness-GEN   shelter
       hoid-is   ***Veeni-ø***   ***käe-ø***   ***alt***   *kinni.*
       hold-PST.3SG   Veeni-GEN   hand-GEN   under.SEP   on to
       'The girl whose face remained in the shelter of darkness held Veeni's arm.'
       [ILU 1990 \ilu0035]

Such instances make up most of the uses that represent the simple structure (1.1 out of 5.5 instances pmw and 4.5 out of 8.5 instances pmw in the 1st and 2nd half of the 20th century respectively). Thus, I assumed it is not the case that *käe all* (hand+under) did not have a non-literal reading in the earlier periods, but rather that these instances are not attested in the corpora.

4.8.2.2. *Käe all* (hand+under) as a complex adverb and
a complex postposition

In contemporary Estonian, *käe all* (hand+under) is predominantly used as a complex postposition. However, it is also possible to use it in an adverbial function (see section 4.4. above). Based on the marginality of the adverbial form of *käe all* (hand+under) in the contemporary data, it can be hypothesized (contra Habicht and Penjam (2007), see section 4.8.1) that the adverbial stage has not preceded the postpositional stage in the case of *käe all* (hand+under). In the following, I will discuss the distribution of complex postpositional and adverbial usages of *käe all* (hand+under) in the diachronic data (Figure 35).
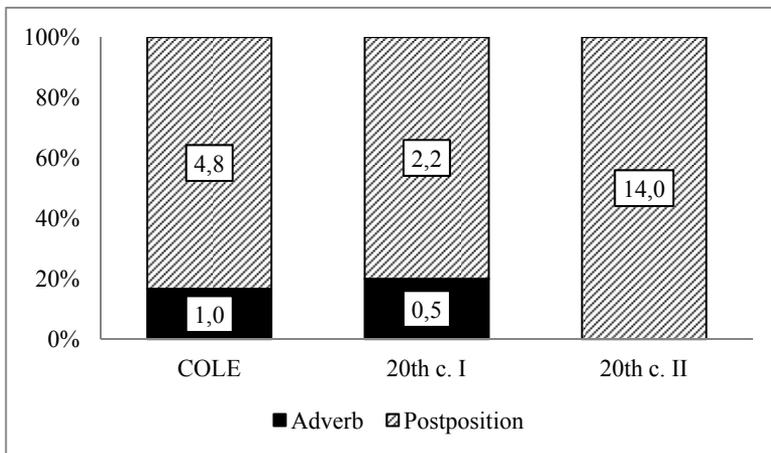
Figure 35. The distribution of the complex postpositions and complex adverbs among usages of *käe all* (hand+under) in the observed periods as instances pmw

The above figure shows that adverbial uses are rather infrequent in the diachronic data as well. However, the available data also suggests that adverbial functions have been more prominent during the earlier periods. In COLE, complex adverbs made up 17% (1.0 out of 5.8 instances pmw) of the usages of *käe all* (hand+under) as a complex item. In the 1st half of the 20th century, the amount remains about the same. However, in the second half of the 20th century, *käe all* (hand+under) does not occur as a complex adverb at all. These results are compatible with the results of the synchronic analysis, where the adverbial uses of *käe all* (hand+under) are rare (see section 4.4). However, as the adverbial function seems to be of greater importance during the earlier periods, a closer analysis of the first examples of *käe all* (hand+under) is needed. Figure 36 depicts the first attestations of *käe all* (hand+under) as a complex adverb and a complex postposition per decade in absolute numbers.
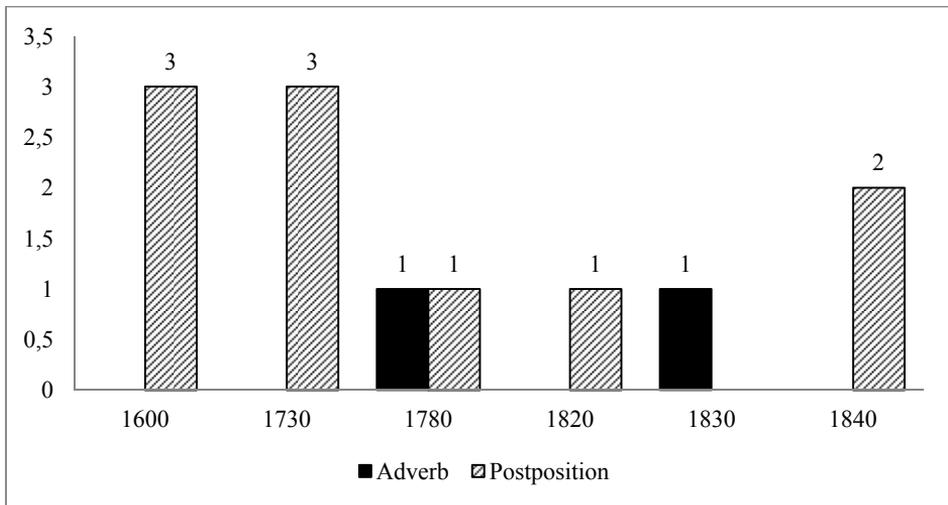
Figure 36. The first attestations of *käe all* (hand+under) as a complex unit in COLE as absolute frequencies per decade

The diachronic data suggests that during the 17th century and the first half of the 18th century, *käe all* (hand+under) is only used as a complex postposition. The available data show that the first complex adverbial usage of *käe all* (hand+under) appears in 1780s (see example (243). This would suggest that *käe all* (hand+under) as a complex unit developed as a complex postposition first, and that the adverbial uses came later, in the end of the 18th century. However, as the examples are – there are only 6 examples of the postpositional use of *käe all* (hand+under) before the first adverbial use – the developmental path of the phrase remains uncertain. Nevertheless, the above suggestion that the hybrid forms might serve as a bridging context to the postpositional usage, supports the claim of postpositional usages preceding adverbial usages. The reason being that the hybrid forms have a similar structure to that of the earliest complex postpositions, i.e. a nominal modifier that refers to God (see examples (238) and (240). So it is possible that the reanalysis as a complex postposition occurred in this context, when the adjectival modifier was dropped.

(243) *Seperräst   olle-te   teije   ka   ärräwanno-tu,   et   te-i-l   kik*
      therefore   be-2PL   you   also curse-PST.PTCP so you-PL-ADE all
      Käe-ø   al   tühja-ø   löppe-p,   sest   teije   pettä-te
      hand-GEN   under.LOC   empty-ILL   end-3SG   because   you   cheat-2PL
      minno-ø   kik.
      I-PRT   all
      Lit. Therefore are you also cursed, that everything under your hand will end in nothing, because you all cheat me.
      'Therefore are you also cursed, so that everything you undertake will come to nothing, because you all cheat me.' [COLE, Frölich, 1787, 13]

210

4.8.2.3. Extension and decatorialization
of the complex postpositional *käe all* (hand+under)

In this section, I discuss the extension and decategorialization of *käe all* (hand+all). Both of them are observed in the features of the (pro)nominal (PN) of *käe all* (hand+under). The PN behaves as a modifier of the body part term *käsi* 'hand' in cases of the simple structure and as a complement of the complex postposition in the complex structure. First, the semantic classes of the PN are observed. The distribution of the semantic classes is presented in Figure 37.
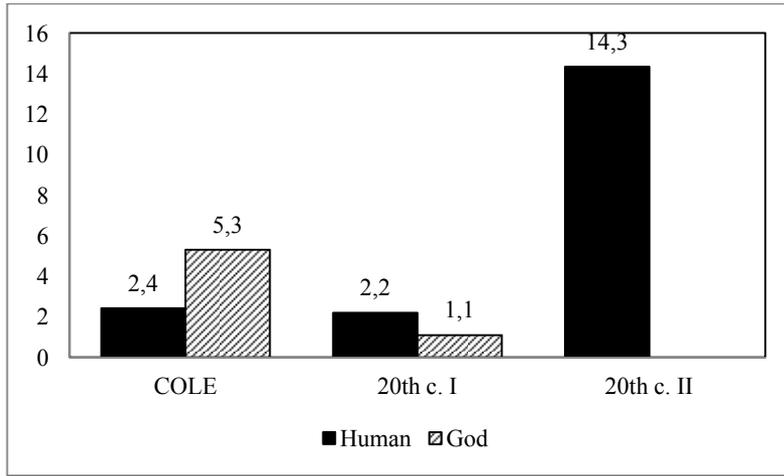
Figure 37. The distribution of the semantic classes of PNs that co-occurred with *käe all* (hand+under) as instances pmw

The data show that *käe all* (hand+under) is only used with PNs that refer to humans and God. It was demonstrated in the synchronic analysis (see section 4.5.1.1), that this is also mostly the case in contemporary Estonian. However, the contemporary data show that *käe all* (hand+under) is also used with PNs that refer to collectives. As such usages only occur with complex postpositions and because they do not occur earlier, it is likely that *käe all* (hand+under) has extended to new contexts in contemporary language. However, the small data sample again casts some doubt on this conclusion.

   As in the synchronic analysis, decategorialization is observed in the non-agreement in the number of the body part term and the PNs (see sections 2.5.3.2 and 4.5.2.1). There were no instances of plural PNs amongst the usages of *käe all* (hand+under) in either COLE or the texts that represent the first half of the 20th century. The first attestations of plural PNs emerge only in the second half of the 20th century. Even then, the plural PN is extremely rare – the absolute number of such examples was three, which makes 1.1 instances pmw. All of the instances of plural PNs occurred with cases where *käe all* expresses MENTAL CONTROL (see example (244)). However, the synchronic analysis suggests that the plural is more common in contemporary language, where the non-

agreement occurs in 21% of all the examples. As the plural PNs also co-occur almost without exception with the complex structure in the contemporary language data, further decategorialization is indicated. However, because the examples are few, this cannot be claimed with certainty.

(244) *Pigem   ajenda-s   min-d   mõte,   miks   ei   või-ks*
rather   motivate-PST.3SG   I-PRT   thought   why   NEG   might-COND
Eesti-s   oll-a   kooli-ø,   kus   ole-ks   võimalik
Estonia-INE   be-INF   school-PRT   where   be-COND   possible
*õppi-da   professionaalse-te   õpetaja-te   **käe-ø   all***
study-INF   professional-PL.GEN   teacher-PL.GEN   hand-GEN   under.LOC
estraadi-ø   ja   varieteezhanri-t.
estrade-PRT   and   variety genre-PRT
Lit. Rather I was motivated by the thought that why couldn't there be a school in Estonia where it would be possible to study estrade and variety genre under the hand of professional teachers.
'Rather I was motivated by the thought that why couldn't there be a school in Estonia where it would be possible to study estrade and variety genre under professional teachers.' [AJAE1990\stak0204]

4.8.2.4. Summary of the diachronic analysis of *käe all* (hand+under)

The first instances of *käe all* (hand+under) appear at the beginning of the 17th century in sermons by Müller. The phrase is used to express non-literal meanings from the very first attestations. The data shows that *käe all* (hand+under) occurs in the diachronic data as a freely combined phrase, a complex unit, and as a hybrid form. Over the observed periods, use of the hybrid forms has decreased, use of the simple structure has increased, and use of the complex structure fluctuates. Nevertheless, it was suggested that the development of *käe all* (hand+under) still adheres to the general principles of grammaticalization. The data indicates that the complex postpositional use might have been affected by the German expression *unter die gewaltige Hand Gottes* 'under the mighty hand of God', because the first examples of the phrase represent such usages. Later, the patterns, where the adjectival modifier is dropped, become more common. This suggests that the development of the complex postpositional *käe all* (hand+under) is affected by loan translations.

The available data suggests that postpositional usages may have preceded adverbial usages of *käe all* (hand+under), because the former appear about 180 years earlier in the data. This path of development is also supported by the fact that the hybrid forms seem to provide a bridging context between the simple and complex structures, and the general marginality of complex adverbs among the usages of *käe all* (hand+under) in the synchronic as well as the diachronic data.

The diachronic data does not include examples with PNs that refer to collectives. As collective PNs are present in the contemporary data, it is possible that it is indicative of extension. Similarly, plural PNs are extremely rare in the dia-

chronic data and do not appear before the 2nd half of the 20th century. As plural attributes are rather common (20%) in contemporary Estonian, it possible that further decategorialization of *käe all* (hand+under) is suggested. However, it is also possible that these differences between the diachronic and contemporary data are due to vast differences between the size and or genres of the corpora used.

4.8.3. *Külje all* (side+under)

The phrase *külje all* (side+under) first appears in the corpus in 1910. The examples from the beginning of the 20th century express the notion of physical proximity to human beings (245) as well as to geographical locations (246), and as such are fairly similar to the use of *külje all* 'side+under' in contemporary language.

(245) *Kui   ta   aga   tund-is,   et   Liisi   tema-ø   kõrwal*
      when   s/he   but   feel-PST.3SG   that   Liisi   s/he-GEN   beside
      maga-s,   lask-is   ta   Liisi-ø   juure,   puge-s
      sleep-PST.3SG   let-PST.3SG   s/he   Liisi-GEN   by   creep-PST.3SG
      *õe-ø   **külje-ø   alla,**   hakka-s   kahe-ø*
      sister-GEN   side-GEN   under.LAT   start-PST.3SG   two-GEN
      käe-ga   ta-ø   kaela-st   kinni   ja   wõdise-s.
      hand-COM   s/he-GEN   neck-ELA   to   and   jiggle-PST.3SG
      Lit. But when s/he felt that Liisi beside him/her slept, s/he lay down by Liisi,
      crept under his/her sister's side, grabbed her neck with two hands and trembled.'
      'But when s/he felt that Liisi beside him/her slept, s/he lay down by Liisi, crept
      by his/her sister's side, held her around her neck with two hands and trembled.'
      [ILU1910\ilu0002]

(246) *Wõru-ø   **külje-ø   alt**   alga-s   uus   wõidukäik*
      Wõru-GEN   side-GEN   under.SEP   start-PST.3SG   new   triumph
      kuni   Marienburi-ø   järwe-ni.
      to   Marienburg-GEN   lake-TER
      Lit. Under the side of Wõru started a new triumph up to lake Marienburg.
      'Near Wõru started a new triumph that lasted up to lake Marienburg.'
      [AJA1910\pm0395]

The total number of examples in the 20th century data is only 45. However, the fact that the phrase does not occur in COLE at all, although its individual components – the body part term *külg* 'side' and the simple postposition *all* 'under' – are present, suggests that it might be possible to cast a glance at the very beginning of the developmental path of *külje all* (side+under) as a complex unit.

    In the following, I present an analysis of *külje all* (side+under) based on the data from the beginning of the 20th century. The findings will be contrasted with that of the contemporary data. The analysis will follow the structure introduced in section 4.8.1.

### 4.8.3.1. Frequency of *külje all* (side+under) and the distribution of the simple and the complex structure

In this section, I discuss the distribution of examples of *külje all* (side+under) in the 1st and 2nd halves of the 20th century in comparison with contemporary data from etTenTen. As the examples of *külje all* (side+under) are similar to contemporary usages, the usages are divided into freely combined phrases and complex units. Firstly, I present the overall frequency of *külje all* (side+under) in Figure 38.
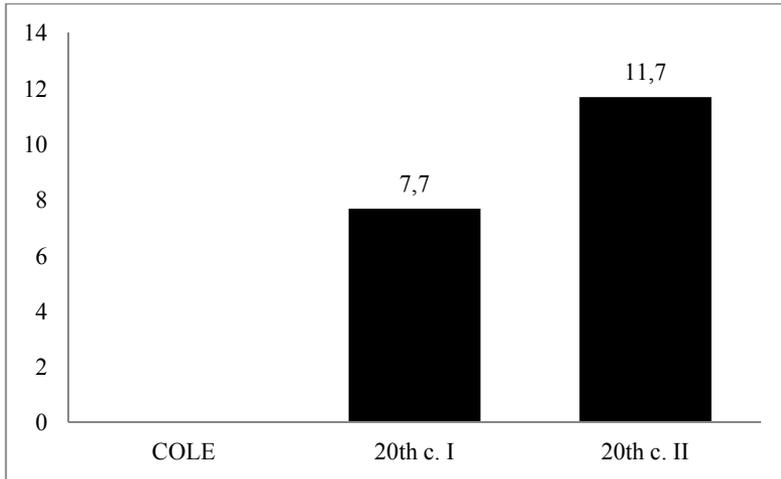
Figure 38. The frequency (pmw) of *külje all* (side+under) in the observed periods

Figure shows that the frequency of *külje all* (side+under) increases during the 20th century (from 7.7 to 11.7 instances pmw). In the contemporary data, the phrase is so frequent that the figures remain of the same magnitude (11.0 instances pmw) even though the sizes of the corpora differ greatly. Thus, it seems that ever since *külje all* (side+under) appeared at the beginning of 20th century, its frequency has been increasing.

In the following, I present the distribution of freely combined phrases and complex units in the diachronic data of *külje all* (side+under) (Figure 39).

214
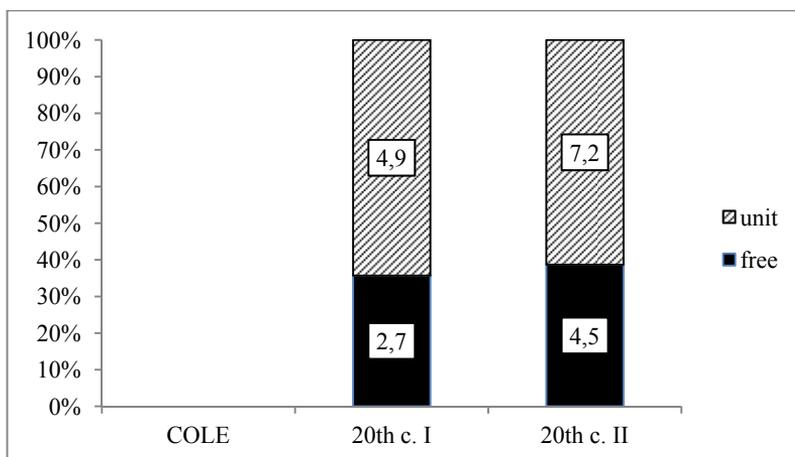
Figure 39. The distribution of *külje all* (side+under) as a freely combined phrase and a complex unit in the observed periods as instances pmw

Figure 39 shows the relative frequencies of *külje all* (side+under) as a freely combined phrase and as a complex unit. At the beginning of the 20th century, *külje all* (side+under) functions as both a freely combined phrase and a complex unit. Thus, it is possible that even though there are no examples in the data from earlier periods, the phrase was already used and even lexicalized. The relative frequency of complex units among all the examples of *külje all* (side+under) remains more or less the same in both halves of the 20th century 64% (4.9 out of 7.6 instances pmw) in 1st half of the 20th century and 61% (7.2 out of 11.7 instances pmw) in the second half of the 20th century. It was also observed that in the contemporary data the proportion of complex units is 86% (see section 4.2). Thus, the available data suggest that the development of *külje all* (side+under) is consistent with the general principles of grammaticalization.

4.8.3.2. *Külje all* (side+under) as a complex adverb and a complex postposition

In contemporary Estonian, *külje all* (side+under) as a complex item may serve as an adverb as well as a postposition. The data suggest that both parts-of-speech also occur in the diachronic data.[67] The distribution of complex adverbs and complex postpositions of *külje all* (side+under) are given in Figure 40.

---

[67]   Although a few (8 out of 2957, less than 1%) instances of hybrid forms appear in the contemporary data of *külje all* (side+under) (see section 4.2), no such instances occur in the diachronic data.
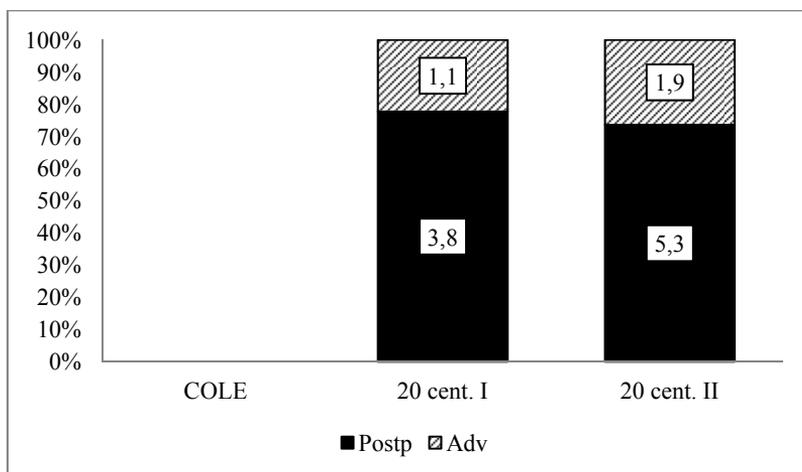
Figure 40. The distribution of the complex postpositions and complex adverbs among the usages of *külje all* (side+under) in the observed periods as frequency pmw

The available data suggest that in case of *külje all* (side+under), postpositional use prevails over adverbial use in both halves of the 20th century. Figure 40 shows that both the adverbial uses as well as the postpositional uses were already present in the 1st part of the 20th century, where the postpositional uses make up 77% of the usages. In the 2nd half of the 20th century, the proportion of postpositional and adverbial uses remains about the same (73%). As shown in section 4.4, the proportion of postpositions in contemporary language is even higher at 90%. Since postpositional uses are regarded as more grammatical than adverbial uses (see section 2.2), it seems that in this aspect the development of *külje all* (side+under) as a complex unit is consistent with the general principles of grammaticalization.

Nevertheless, the data does not offer a clear answer regarding the diachronic order of the parts-of-speech in the developmental path of *külje all* (side+under) as a complex unit. Figure 41 gives the absolute frequencies of each part-of-speech per decade.
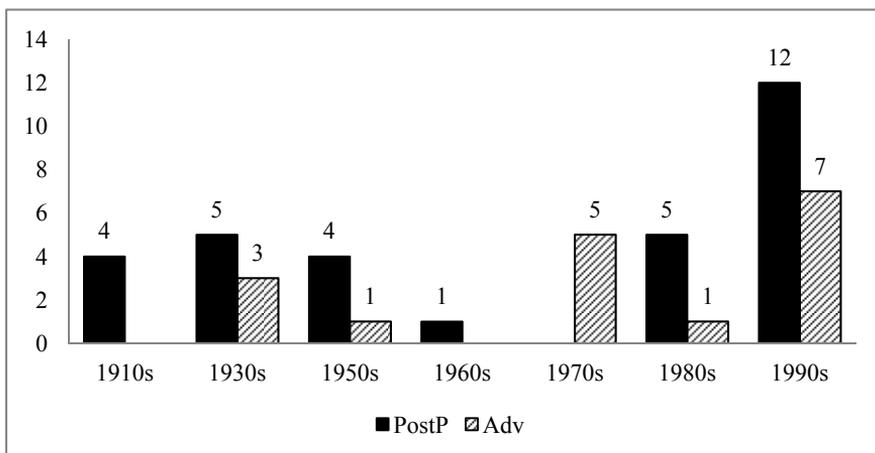
Figure 41. The first attestations of *külje all* (side+under) as a complex unit in the 20th century as absolute frequencies per decade

It can be observed that the first postpositional use ((246) in section 4.8.3) occurs in the 1910s, and the first adverbial use (247) in the 1930s.

(247)   *Ei,   tõesti,   ei   ole-ø   sugugi   paha,   et   nii*
        no   really   NEG   be-CONNEG   at all   bad   that   so
        imeliku-l   viisi-l   sattu-s   ela-ma   just   siia
        weird-ADE   way-ADE   happen-PST.3SG   live-SUP   exactly here
        külje-ø   alla.
        side-GEN   under.LAT
        Lit. No, really, it is not at all bad that in such a weird way happened to live here under the side.
        'No, really, it is not at all bad that in such a weird way [s/he] happened to come to live near here.' [ILU1937\ram0039]

Although complex postpositional uses appear before complex adverbial uses, the distance between the two occurrences in only twenty years. Even though the postpositional uses are prevalent in all the decades except for the 1970s, it cannot be concluded the data reflects the actual course of development of *külje all* (side+under), because of the extremely small number of available examples. Thus, in the case of *külje all* (side+under), the order of the appearance of the adverbial and postpositional functions remains inconclusive.

4.8.3.3. Extension of the complex postpositional *külje all* (side+under)

Following the logic of synchronic analysis, the extension of *külje all* (side+under) is observed in the semantic class of the (pro)noun that precedes *külje all* (side+under).

6

5

4

3

2

1

0

COLE   20th c. I   20th c. II

0,5   3,3   1,6   1,1   0,8   4,9

■ Animate   ▨ Artificial object   ▤ Region
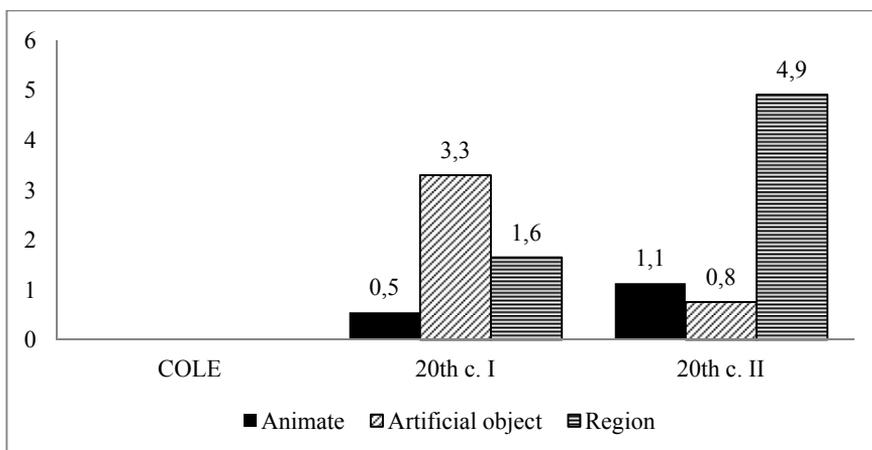
Figure 42. The distribution of the semantic classes of the PNs that co-occur with *külje all* (side+under) as instances pmw

The figure shows that the use of *külje all* (side+under) is rather diverse in the 1st half of the 20th century, when it first appears in the diachronic data. The data suggest that during this period, the phrase began to be used with PNs that referred to animate beings, artificial objects and regions. In the data, the animate PNs only refer to humans. The same classes are used in the 2nd half of the 20th century. However, contemporary data includes two more classes not represented in the earlier periods – natural object and collective/abstract (see section 4.5.1.2). It is possible that these did not occur in the earlier periods because of a lack of data, but it may also be the case that *külje all* (side+under) did not co-occur with such nouns during the earlier periods. This would, especially in case of collective/abstract PNs, suggest extension because as mentioned in section 4.5.1.2, collective and abstract PNs only co-occur with instances where *külje all* (side+under) is used as a complex unit.

    Figure 42 also shows that the relative frequency of PNs that refer to regions increases over the investigated periods. The relative frequency of cases where *külje all* (side+under) is preceded by a noun referring to region increases from 1.6 instances (pmw) in the 1st half of the 20th century to 7.7 instances pmw in contemporary data. Region is the most common type of PN starting from the 2nd half of the 20th century. As was shown in the synchronic analysis (see section 4.5.1.2), this is the most common (81%) semantic class among the PNs of *külje all* (side+under).

    Figure 43 shows that the PNs that refer to regions (246) occur only with complex units only, and that they make up the majority of the PNs that occur with complex uses in the 20th century. Thus, the overall increase of instances of complex units is mainly due to increases in this kind of example.
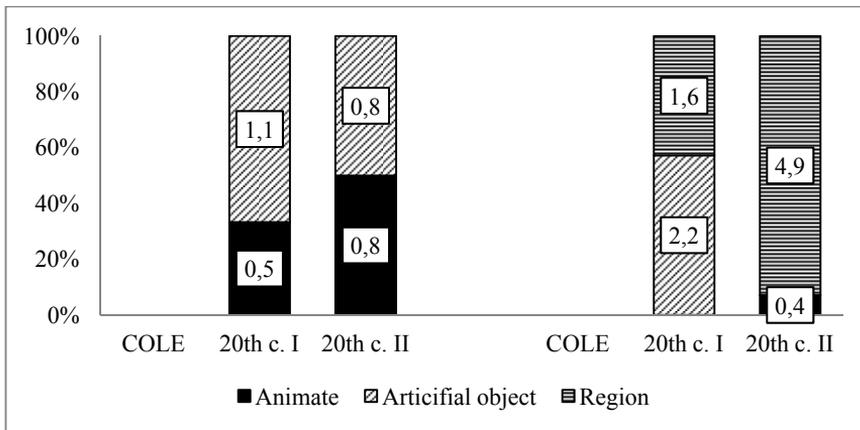
Figure 43. The distribution of the semantic classes of PNs of *külje all* (side+under) when used as a freely combined phrase and a complex unit as instances pmw

The data also suggests that animate PNs are possible with the simple structure (248), as well as the complex structure (249), and that PNs referring to objects are also possible with both structures (see examples (250) and (251)). As was shown in section 4.5.1.2, the synchronic analysis yields similar results. However, in contemporary language, the complex postpositional *külje all* (side+under) may also be complemented by a noun that refers to collectives or abstract notions. Thus, based on the available data, it seems that PNs that refer to regions suggest an extension of *külje all* (side+under), because they do not appear in the simple structure, and collective/abstract PNs suggest extension, because they do not appear with the simple structure and seem to be a more recent development. However, it may be due to the lack of available data.

(248) *Toidupoolis,   mis   Hargula-st   an-t-i ,   jä-i*
      food   that   Hargula-ELA   give-IMPS-PST   leave-PST.3SG
      Aivari-ø   külje-ø   alla   vereloiku-ø.
      Aivar-GEN   side-GEN   under.LAT   blood puddle-ILL
      'The food that was given in Hargula was left under Aivar's side in a puddle of blood.' [ILU1980/ tkt0109]

(249) *Kui   Eva   vahel   harva   tusatse-b ,   poe-b   ta   rohkem*
      if   Eva   sometimes rarely sulk-3SG   creep-3SG   s/he   more
      Kaarli-ø   külje-ø   alla;   mu-lle   kuulu-b   naise-ø
      Kaarel-GEN   side-GEN   under.LAT   I-ALL   belong-3SG   woman-GEN
      päikeseline   pool.
      sunny   side
      Lit. If Eva sometimes sulks, she creeps more under Kaarel's side; to me belongs the woman's sunny side.
      'When Eva sometimes sulks, she creeps closer to Kaarel; I have the woman's sunny side.' [ILU1980/ tkt0110]

219

(250) *Siin-seal*   *küll*   **_mõne-ø_**   **_kivi-ø_**   **_külje-ø_**   **_all_**
here and there   indeed   some-GEN   stone-GEN   side-GEN   under.LOC
viletsalt   *valge-t,*   *paar*   *peotäi-t*   *vahest.*
badly   white-PRT   couple   handful-PRT   maybe
'Here and there, under the side of some stone, a measly amount of white, maybe a couple of handfuls.' [ILU1936/ ram0020]

(251) *Sest*   *vana*   *Juurup*   *sa-i*   *väga*   *hästi*   *aru-Ø,*
because   old   Juurup   get-PST.3SG   very   well   sense-PRT
et   *Kaarli-l*   *mujale*   *töö-d*   *otsi-ma*   *minn-a*
that   Kaarel-ADE   elsewhere   work-PRT   find-SUP   go-INF
p-ol-nud   *kerge ;*   *Kaarli-ø*   *onn,*   *perekond*   *ja*   *maalapp*
NEG-be-PST.PTCP   easy   Kaarel-GEN   hut   family   and   plot
ol-i-d   *ju*   **_Juusa-ø_**   **_talu-ø_**   **_külje-ø_**   **_all_**.
be-PST-3PL   well   Juusa-GEN   farm-GEN   side-GEN   under.LOC
Lit. Because old Juurup understood very well that it wasn't easy for Kaarel to go elsewhere to look for work; Kaarel's hut, family and plot of land were under the side of Juusa farm.'
'Because old Juurup understood very well that it wasn't easy for Kaarel to go elsewhere to look for work; Kaarel's hut, family and plot of land were near Juusa farm.' [ILU1950/ ilu0017]

Based on the contemporary data (see sections 4.5.1.2 and 4.5.1.7), the analysis of the developmental path of the complex unit of *külje all* (side+under) remains inconclusive. It was unclear whether the complex unit that expresses PROXIMITY developed in the context of PNs that refer to humans or objects. It was suggested that as the body part term *külg* 'side' is productively used with both animate beings and objects, it is possible that the complex item developed in both these contexts simultaneously. There is not much to be added to this discussion based on the diachronic data. As both animate and object PNs are used with both structures, both paths can be considered possible. Thus, the development from usages as exemplified in (248) are possible sources for example (249) and usages as exemplified in (250) are possible sources for examples (251) and (246). However, no such order can be concluded based on the available diachronic data, because example (246) is one of the earliest available examples. Moreover, the data suggests that PNs that refer to objects appear with the complex structure earlier (1st half of the 20th century) than the animate PNs (2nd half of the 20th century). However, as there is so little available data, it cannot be taken as suggestive of earlier attestations.

In conclusion, the data suggest that the semantic classes of *külje all* (side+under) have become more diverse, because in the contemporary data, the phrase is additionally used with PNs that refer to collectives/abstract entities and natural objects. However, the uses of *külje all* (side+under) may be more restricted in the earlier periods owing the small amount of data available. Based on extant examples, both of the developmental paths of *külje all* (side+under) suggested in 4.5.1.7 are possible.

4.8.3.4. Decategorialization of the complex postpositional
külje all (side+under)

In this section, I discuss the decategorialization of the complex postpositional *külje all* (side+under). Similarly to the synchronic analysis, decategorialization is observed in the agreement in number between the body part *külg* 'side' and the preceding (pro)noun (PN). As stated above (cf. 2.5.3.2), the non-agreement are taken to suggest that the body part phrase is interpreted as a complex unit, and that the preceding nominal is analyzed as the complement of the whole utterance. The distribution of singular and plural PNs of *külje all* (side+under) is depicted in Figure 44.
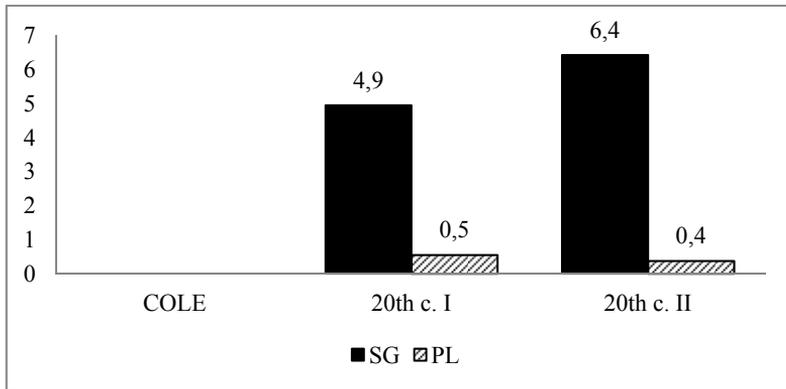
Figure 44. The distribution of singular and plural PNs of *külje all* (side+under) in the observed periods as frequency pmw

Figure 44 suggests that plural PNs used with *külje all* (side+under) first appeared in the 1st half of the 20th century. However, the relative frequency of plural PNs remains extremely low throughout the 20th century. In the 1st half of the 20th century, the frequency of plural PNs is 0.5 instances pmw out of 5.4 (9%) and in the 2nd half of the 20th century 0.4 instances pmw out of 6.8 (6%), which means that there is only one such example in each period. A similar situation was observed in the contemporary data, where plural forms made up only 3% of the examples (see section 4.5.2.2).

A closer analysis of the examples with the plural PNs reveals that one of these instances appears in an example where *külje all* (side+under) is analyzed as a freely combined phrase (252), and the other one occurs with a complex unit (253). In the contemporary data, plural PNs also occur with both structures.

(252) *Varju-d*   **_kadaka-te_**   **_ja_**   **_kivi-de_**   **_külje-ø_**
shadow-PL   juniper-PL.GEN   and   stone-PL.GEN   side-GEN
all   *on*   *üsna*   *lühikese-d*   *ja*   *nõrga-d* – *jõuluvalgus*
under.LOC be.3PL   quite   short-PL   and   weak-PL   Christmas light
voola-b   *ju*   *otse*   *alla*   *maa-le.*
flow-3SG   well   straight   under   land-ALL
'The shadows under the side of junipers and stones are quite short and weak –
Christmas light flows straight down to ground.' [ILU1936\ram0020]

(253) *Mujal,*   *näiteks*   *Ungari-s,*   *on*   *ringraja-d*
elsewhere for example Hungary-INE be.3PL raceway-PL
tähtsa-ma-te   **_keskus-te_**   **_külje-ø_**   **_all_**.
important-COMP-PL.GEN   centre-PL.GEN   side-GEN   under.LOC
Lit. Elsewhere, for example in Hungary, the raceways are by the side of
important centers.
'Elsewhere, for example in Hungary, the raceways are near important centers.'
[AJA1990\ee1443]

Due to the extreme marginality of the plural forms, it is difficult to conclude anything but that plural PNs are extremely rare, though not impossible with instances of *külje all* (side+under), and that the data suggest there has been no change in their frequency.

4.8.3.5. Summary of the diachronic analysis of *külje all* (side+under)

The first attestation of *külje all* (side+under) originates from 1910. Thus, the data suggest that *külje all* (side+under) was not used in the period of Old Literary Estonian, neither as a complex unit nor as a freely combined unit. However, in the first available examples at the beginning of the 20th century, *külje all* (side+under) is used as a freely combined phrase as well as a complex item. Thus, it may be the case that similarly to most of the studied phrases, the data does not reveal the very beginning of the developmental path of *külje all* (side+under). Nevertheless, the available data suggest that the general frequency of *külje all* (side+under), as well as the proportion of uses as a complex unit, have increased. This is consistent with the principles of grammaticalization.

    The diachronic data shows that postpositional uses are already occurring at the beginning of the 20th century, and throughout the 20th century *külje all* (side+under) is used as a complex postposition more and more. However, as the first attestations of adverbial and postpositional uses are chronologically just a few decades apart, it remains unclear which one appeared first.

    The data suggest that in terms of the semantic classes, the PNs were rather diverse already at the beginning of the investigated period (17th to 20th century). Both the simple and complex structures are used with PNs that refer to animate beings and artificial objects. Regions are, again only used to complement complex units. These results are compatible with that of synchronic data. However,

in contemporary language the complex postpositional *külje all* (side+under) may also be used with nouns that refer to collectives or abstract notions. The fact that no such examples were present in the diachronic data may suggest an extension of *külje all* (side+under), though more historical data would be required to confirm this conclusion.

As there are very few available examples, no additional information could be found on the possible developmental path of the complex postpositional use of *külje all* (side+under). Based on the synchronic analysis, it was hypothesized that the complex function words used to express PROXIMITY may have developed in the context of animate PNs or object PNs. Based on the diachronic analysis, neither of these paths can be refuted. Of course, it is possible that this function developed in both of these contexts simultaneously. The data suggest that it is possible to use *külje all* (side+under) with a plural PN, but such examples are very rare at the beginning of the 20th century as well as in contemporary language. The available data suggest that there is no increase in plural PNs used with *külje all* (side+under) from the 17th to the end of the 20th century.

4.8.4. *Selja taga* (back+behind)

The first instances of *selja taga* (back+behind) appear in the corpus data in 1605. Similarly to the phrase *käe all* (hand+under), *selja taga* is first attested in Müller's sermons. In its first attestation, the phrase expresses a literal meaning, i.e. the relation between two physical objects – a father's position relative to his child (see example (254)).

(254)   *Nüith*   *eb*   *pane-ø*   *v̈x*   *Issa*   *oma-ø*   *Nohre-ø*
       now   NEG   put-CONNEG   one   father   own-GEN   young-GEN
       Lapsukeße-ø   *pæle*   *enamb*   *mitte,*   *kudt*   *tæma* *harrenda-b*
       kid-GEN   on   anymore   not   but   s/he   be able-3SG
       kand-a,   *waidt* *sÿßkit*   *keu-b*   *se*   *Issa*   *oma-ø*
       carry-INF   only   however   walk-3SG   this   father   own-GEN
       Lapse-ø   *korwal,* *echk*   <u>*tæma-ø*</u>   ***Selia-ø***   ***tacka,***   *hoÿa-b,*
       kid-GEN   beside   or   s/he-GEN   back-GEN   behind.LOC   hold-3SG
       echk   *awita-b*   *kz*   *kand-a,*   *eth*   *eb*   *se* *Laps* *v̈chteki-t*
       perhaps   help-3SG   also   bear-INF   that   NEG   this   kid   single-PRT
       kachio-ø   *pidda-ø*   *sa-ma.*
       damage-PRT   must-CONNEG   get-SUP
       'Now a father does not put more on his young child than s/he can carry but
       nevertheless the father walks by his child or behind his/her back so that the child
       will not come to any harm.' [COLE, Müller, 1605, s. 29, 5]

The phrase *selja taga* (back+behind) occurred in the diachronic corpora on 494 occasions. Thus, it is over five times more frequent than the phrase *käe all* (hand+under) (96 examples). However, due to scarcity of data in during the earlier periods as well as the genre differences between the corpora (see section

3.2.2), the results of the diachronic analysis of *selja taga* (back+behind) should be taken with caution. In the following, I present an analysis of the diachronic data of *selja taga* (back+behind), which follows the structure introduced in section 4.8.1.

4.8.4.1. Frequency of *selja taga* (back+behind) and the distribution of the simple and the complex structure

Figure 45 shows that the total number of all instances of the phrase *selja taga* (back+behind) increases from the 17th–19th to the second half of the 20th century. In COLE, the total number of occurrences of the phrase amounts to 10.6 instances pmw, by the 1st half of the 20th century it has risen to 81.2 instances pmw, and by the 2nd half of the 20th century, to 122.3 instances pmw.
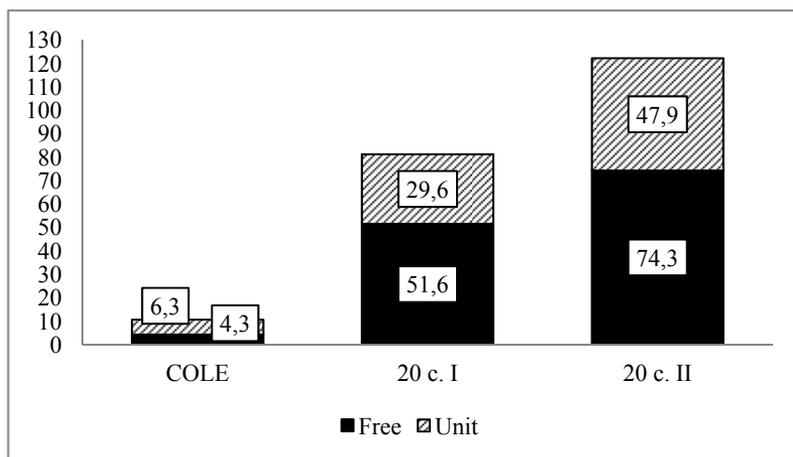
Figure 45. The distribution of the freely combined phrases and complex units of *selja taga* (back+behind) in the observed periods as instances pmw

Figure 45 also shows the relative frequencies of the freely combined phrases and complex units. Both these structures are already present in COLE. The data suggest that the structures were more or less equally frequent in COLE, but in the 20th century the simple structure is a little more frequent. It can be observed in the figure that in COLE, the simple structure occurs on 4.3 instances (pmw) and the complex structure on 6.3 instances. As the overall frequency of the phrase increases throughout the observed periods, the relative frequencies of free as well as the complex uses increase too. In both halves of the 20th century, the simple structure is a little more frequent (51.6 and 74.3 out of 81.2 and 122.2 uses pmw respectively). Therefore, the data suggest that while the overall frequency of *selja taga* (back+behind) has increased, the proportions of the simple and the complex structure has remained more or less the same. However, if contrasted with the results of the synchronic analysis, where the simple

structure makes up 45% of the uses of *selja taga* (back+behind) (see section 4.3.3), it could be suggested that the proportion of the complex units is slowly but steadily increasing.

As *selja taga* (back+behind) carries many unrelated functions (see section 4.3.3), the distribution of these functions is presented separately. Figure 46 shows the relative frequencies of each function during the observed periods, as well as the proportions of the functions within each period.
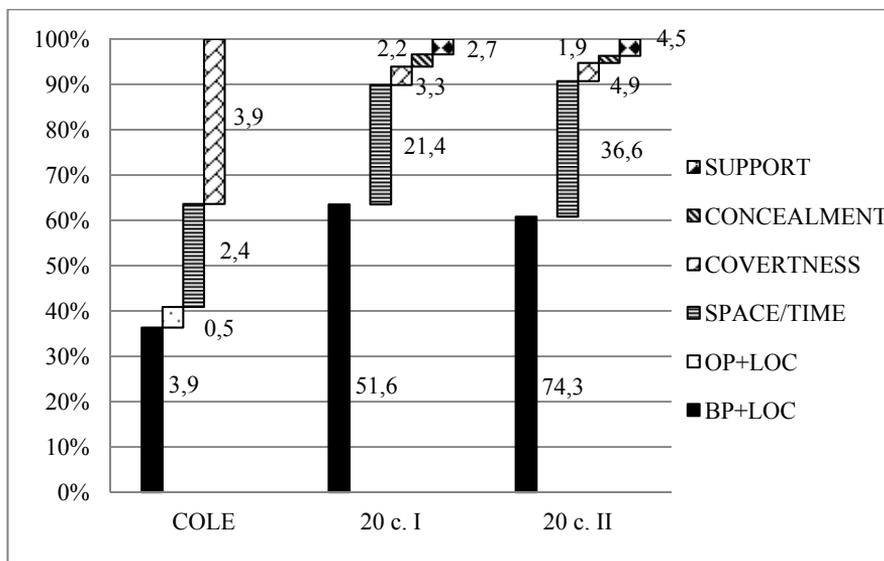
Figure 46. The distribution of the functions of *selja taga* (back+behind) in the observed periods, instances pmw

When observed separately, the functions show a slight increase in complex uses during the 20th century, mainly as a consequence of increased use of the spatio-temporal function (for an example see (255)). The spatio-temporal function makes up 23% (2.4 out of 10.6 instances pmw) of the uses of *selja taga* (back+behind) in COLE, 29% (23.2 out of 81.2) in the 1st half of the 20th century, and 30% (36.6 out of 122.3) in the second half of the 20th century. Again, this is not a large increase, but when contrasted with the contemporary data where the spatio-temporal function makes up 38% of all uses, a tendency of an increase over time can be seen.

(255) *Otsego   üks   suur   unnenäggo   seisa-b   me-i-l   möda*
    as   one   big   dream   stand-3SG   we-PL-ADE   by
    *läi-nud   aeg   **selja-ø**   **tagga**,   ja   uddopilwe-ø*
    go-PST.PTCP   time   back-GEN   behind.LOC   and   fog cloud-GEN
    al   otsego   üks   tundmata   wõeras-maa   meie-ø   eddespiddine
    under   as   one   unknown   foreign land   we-GEN   future
    ello   meie-ø   ees   warjul.
    life   we-GEN   ahead   hidden

    Lit. Like a big dream the bygone time stands behind our back and under a cloud
    of fog like an unknown foreign land our future life hidden ahead of us.
    'Like a big dream the bygone time stands behind us and under a cloud of fog like
    an unknown foreign land our future life hidden ahead of us.'
    [COLE, Kreutzwald, 1843, 3]

While the spatio-temporal function unquestionably makes up the largest amount of examples in the 20th century, the data suggest that in COLE the function COVERTNESS (as exemplified in (256)) was the most prominent among the complex structure. Figure 46 shows that this function makes up 36% of the usages in COLE. However, because the total number of examples in COLE is very small (3.9 out of 10.6 instances pmw), the large proportion may be incidental. This is supported by the fact that COVERTNESS remains rather marginal in both halves of the 20th century (3.3 and 4.9 instances pmw, both 4% of usages), as well as in contemporary language (8%). However, it must be noted that it is still a clearly a distinct function which was already present in Old Literary Estonian.

(256) *Walle-t/   salla-de   Asja-de   jerrelütlemisse-ø   nink   keik*
    lie-PRT   secret-PL.GEN   thing-PL.GEN   repetition-GEN   and   all
    *KeelePexmisse-ø/   **Selja-ø**   **tagga**   kurja-de   Könne-de*
    gossip-GEN   back-GEN   behind.LOC   evel-PL.GEN   talk-PL.GEN
    Töstmisse-ø   nink   Laidmisse-ø.
    lift-GEN   and   blame-GEN

    'Lies, repeating of secret things and all the gossip, speaking evil behind one's
    back and blaming.' [COLE, Blume, 1662, 37]

The functions CONCEALMENT and SUPPORT (see examples (257) and (258) respectively) however do not appear in the corpora before the first half of the 20th century. However, it must be noted that already in the earliest example to be found in the data (example (254)), the function to express SUPPORT is already present. This example originates from a sermon, which include religious rhetoric, i.e. the story of father and child meant to illustrate the relationship between man and God. So in this example the phrase *selja taga* (back+behind) (as a part of a larger unit) can be understood as expressing more than just a physical relation, that is a more abstract relation used to describe support provided by God. However, it is not claimed that *selja taga* (back+behind) as an individual unit carries an abstract meaning in this example, but rather that a

sense of support is here provided by the context. Examples such as this may have served as suitable semantic contexts to show where a more abstract meaning has evolved. However, it is not clear whether the functions SUPPORT and CONCEALMENT were not present during the period of COLE, or that they just do not appear in the corpus. The data suggest that uses of both remain relatively infrequent during the 20th century – CONCEALMENT makes up only 1–2% of the usages of *selja taga* (back+behind) in each period (2.3 of 81.2 and 1.9 of 122.3 instances pmw in the 1st and 2nd half of the 20th century respectively); the function SUPPORT is also fairly marginally used, amounting to around 3% of the usages in each period (2.7 out of 81.2 and 4.5 out of 122.3). In the contemporary data, CONCEALMENT and SUPPORT make up 3% and 6% of all the uses of *selja taga* (back+behind) respectively (see section 4.3.3).

(257) …*Kas*   *koolmeister*   *mõni*   *saks*   *on,*   *et*   *ise*   *oma-ø*
Does   schoolmaster   some   squire   be.3SG   that   himself   own-GEN
ahju-ø   *pragu-ø*   *savi-ga*   *ei*   *või-ø*   *kinni*
oven-GEN   crack-PRT   clay-COM   NEG   could-CONNEG   to
määri-da;   *ela-b,*   *nagu*   *suure-ø*   *Jumala-ø*   ***selja-ø***
loam-INF   live-3SG   like   great-GEN   God-GEN   back.GEN
taga.
behind.LOC
Lit. Is the schoolmaster some squire that he cannot patch the crack in his oven with clay himself; lives as if behind the great God's back.
'Is the schoolmaster some squire that he cannot patch the crack in his oven with clay himself; he lives a convenient and carefree life.' [AJA1900\aja0001]

(258) *Ja*   *see*   *on*   *ka*   *wäga*   *arusaadaw:*   *tema-ø*
and   this   be.3SG   also   very   understandable   s/he-GEN
seljataga   *seis-is*   *mõisaherra*   *oma-ø*   *politseilise-ø*
back.behind.LOC   stand-PST.3SG   manor squire   own-GEN   police-GEN
ja   *koduse-ø*   *karistuslise-ø*   *wõimu-ga.*
and   homely-GEN   punishing-GEN   power-COM
Lit. And this is also very understandable: behind his back stood the manor squire with his power of police and domestic punishment.
'And this is also very understandable: behind him stood the manor squire with his power of police and domestic punishment.' [AJA1900\aja0096]

The diachronic data suggest that the usages of *selja taga* (back+behind) as freely combined phrases are mostly combinations of the body part term and a simple locative function word (BP+LOC). Instances where *selja taga* (back+behind) is a combination of an object part and locative function word (OP+LOC) are extremely rare. Nevertheless, such a usage was already present in the first diachronic period, in fact, the first and only such example originates from 1849 (see example (259)).

(259) *Kui*   *se*   *tö*   *walmis,* *siis* *minne-ø* <u>*pu-ø*</u>   **selja-ø**
       when this work done then go-IMP beehive-GEN back-GEN
       tahha *kula-ma,* *kas* *seäl* *sees* *weel* *heäl-t*
       behind.LAT listen-SUP if there inside more voice-PRT
       kule-d.
       hear-2SG
       Lit. When this work is done then go behind the beehive's back to listen if you
       can still hear a voice in there.
       'When this work is done then go behind the beehive to listen if you can still hear
       a voice in there.' [COLE, Freundlich, 1849, 86]

In this example, the phrase is used to refer to a location relative to an artificial object (*pu selja taha* LIT. 'behind the back of a beehive'). Although examples with object parts do not appear in the data from the 20th century, they do occur as a relatively marginal feature in the etTenTen (<1%). This suggests that usages with object parts existed throughout the 17th to 20th century, but did not end up in the latter corpora due to their rareness. This suggests that the object part is an unlikely gateway for the development of the spatio-temporal complex function word *selja taga* (back+behind), as would be expected based on the previous research (e.g. Svorou 1994: 90; Heine 1997: 44). I also reached the same conclusion based on the synchronic analysis of *selja taga* (back+behind) (see section 4.5.1.7). Of course, the scenario that the development of the spatio-temporal function had been influenced by the object part function before the 17th century in spoken language cannot be completely excluded. However, it is not considered very likely because almost all of the examples of this function in the diachronic corpora occur with human PNs (see examples (260)–(262)).

However, a closer look at the first attestation of the spatio-temporal function does not shed much light on its development. As stated in section 4.3.3, the spatio-temporal function includes three functions that are not always separable – locative, temporal and ordinal. The diachronic data suggest that all of these functions first appeared in the data during the 19th century and over a relatively short period of time. The first attestation of the spatio-temporal *selja taga* (back+behind) originates from the 1840s and is used to express a temporal relation (see example (260)); the first example used to express a locative meaning appeared just two decades later in the 1860s (see example (261)); the first example used to express an ordinal meaning originates from the 1890s (see example (262).

(260)   *Kui   meie   Jummala riki-ø   ja   temma-ø   öigus-t*
         if   we   kingdom of God-PRT   and   s/he-GEN   justice-PRT
         noua-me,   siis   tarwita-b   se   kül,   et   meie   keik,   ja   ka
         instist-1PL   then   use-2SG   this   well   that   we   all   and   also
         se-dda,   mis   me-i-l   siin   ilma-s   keigearmsa-m
         this-PRT   what   we-PL-ADE   here   world-INE   loveliest-COMP
         on,   selja-tahha   heida-me,   kui   meie   se-dda
         be.3SG   back-behind.LAT   through-1PL   when   we   this-PRT
         tunne-me,   et   se   me-i-d   kela-b   selle-ø
         feel-1PL   that   this   we-PL-PRT   disallow-3SG   this-GEN
         etteseatud   märgi-ø   pole   minne-ma-st.
         being ahead   sign-GEN   toward   go-SUP-ELA
         Lit. If we insist on the kingdom of God and His justice then it is necessary that
         we leave everything behind our back, including that which is the most dear to us
         in this world, if we feel that it prevents us from striving towards that set goal.
         'If we insist on the kingdom of God and His justice then it is necessary that we
         leave everything behind, including that which is the most dear to us in this
         world, if we feel that it prevents us from striving towards that set goal.' [COLE,
         Kersten, 1847, 71]

(261)   *Igga   lille-le,   igga   puwõssa-le   kaeba-s   ta*
         every   flower-ALL   every   tree bosk-ALL   complain-PST.3SG   s/he
         nutt-es   omma-ø   önnetus-t   ja   kurwastus-t, –   kui   nee-d
         cry-GER   own-GEN   calamity-PRT   and   sadness-PRT   if   this-PL
         ta-ø   selja-ø   tahha   jä-nud;   siis   oll-i
         s/he-GEN   back-GEN   behind.LAT   stay-PST.PTCP   then   be-PST.3SG
         ta   kui   hea   söbra-de-st   lahku-nud;   sest
         s/he   as   good   friend-PL-ELA   leave-PST.PTCP   because
         muu-d   söbra-ø   temma-l   ennam   ei   ol-nud,   kui
         else-PRT   friend-PRT   s/he-ADE   anymore   NEG   be-PST.PTCP   as
         Jummala-ø   rohho-ø   taime-d.
         god-GEN   cure-GEN   plant-PL
         Lit. To every flower, every tree bosk s/he lamented, crying, his/her calamity and
         sadness – when these were left behind his/her back it was as if s/he had left
         good friends because s/he no longer had other friends besides God's plants.
         'To every flower, every tree bosk s/he lamented, crying, his/her calamity and
         sadness – when these were left behind him/her it was as if s/he had left good
         friends because s/he no longer had other friends besides God's plants.' [COLE,
         Mai Roos, 1865, 175]

(262) *Libeda-ø   keele-ga   nais-i   on   kõig-i*
      slick-GEN   tongue-COM woman-PL.PRT   be.3PL all-PL.GEN
      rahwas-te   seas,   kuid   Maori-ø   naese-d
      nation-PL.GEN   among   but   Maor-GEN   woman-PL
      on   palju   kõnerikkama-d   ja   jäta-wad   selle-ø   poolest
      be.3SG   more   talkative-PL   and   leave-3PL   this-GEN   by
      *kõik   teise-d   kaugele   **selja-ø   taha**.*
      all   other-PL   far   back-GEN   behind.LAT
      Lit. There are women with slick tongues among all nations but Maori women are
      much more talkative and leave all others far behind their backs in this.
      'There are women with slick tongues among all nations but Maori women are
      much more talkative and exceed all others in this.' [AJA1890\pro0082]

As these functions appear more or less at the same time, there is no solid evidence to propose any diachronic order of their development. However, given its frequency of use in the contemporary and diachronic data, it is possible that the temporal preceded the locative and the positional. This scenario is also supported by the semantic properties – the temporal is much more conventionalized as a separate function, i.e. it is clearly different from BP+LOC because it expresses a meaning of another, more abstract domain (time). The locative meaning however, belongs to the same domain (space) as the source form (BP+LOC) and therefore these two are practically indistinguishable. Thus, when the temporal meaning 'passed in time' arose, it is possible that it also supported the autonomization of the meaning 'passed in space'. The positional function is closer to the latter, which could mean the complex locative meaning is its predecessor. However, the ordinal *selja taga* (back+behind) might also be influenced by the source form (BP+LOC), as these usages are semantically very close and in actual use are not distinguishable in every case. However, the rest of the functions – COVERTNESS, CONCEALMENT and SUPPORT – are distinguishable. As there are no bridging contexts in the diachronic data that would indicate that these functions belong to the same developmental path, they are viewed as separate branches of grammaticalization of the source form and thus where necessary are discussed separately.

4.8.4.2. *Selja taga* (back+behind) as a complex adverb and a complex postposition

Similar to the phrase *käe all* (hand+under), *selja taga* (back+behind) may as a complex unit function as both the complex adverb and complex postposition. As discussed in section 4.4, in contemporary Estonian *selja taga* (back+behind) is used frequently in both functions, but mostly as a complex adverb in the spatio-temporal function. In the following, I will discuss the distribution of complex adverbs and postpositions among the complex units in the diachronic data.
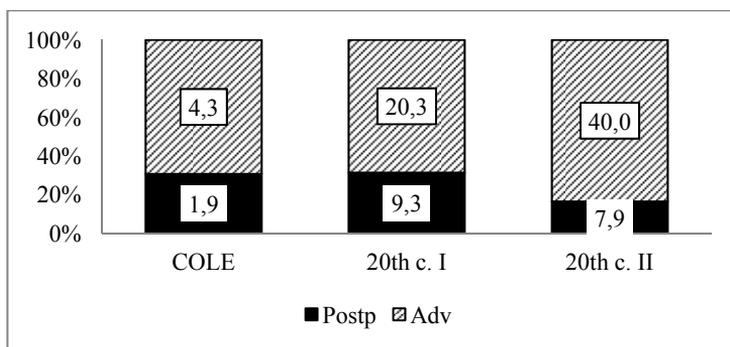
Figure 47. The distribution of the complex postpositions and complex adverbs among the usages of *selja taga* (back+behind) in the observed periods as instances pmw

Figure 47 shows the frequencies (pmw) of *selja taga* (back+behind) when used as a complex postposition and as a complex adverb, as well as the proportions of each part of speech among all the instances where the phrase is used as complex unit. The data suggest that in all of the observed periods, *selja taga* (back+behind) is more frequently used as a complex adverb than as a post-position. Although the absolute number of postpositional uses increases in the 1st half of the 20th century compared to the 17th–19th centuries, the proportion of the postpositions remains the same (around 30%) in these two periods. By the 2nd half of the 20th century, the number of postpositional uses decreases a little (to 7.9 instances pmw), and the proportion of postpositional uses drops below 20% of all the uses of *selja taga* (back+behind) as a complex unit. This is in line with the results of the synchronic analysis, where it was shown that adverbial uses make up 75% of all the uses of *selja taga* (back+behind) as a complex unit.

Thus, although the diachronic data show that *selja taga* (back+behind) is used increasingly more in general, and that its usage as a complex item are slowly increasing, the data does not indicate an increase of postpositional usages at the expense of adverbial usages over time. Such a result would be expected because the postpositional usages are considered more grammatical than the adverbial usages. Of course, such results can be explained by the fact that the studied phenomenon is in its early stages. It can be expected that the expansion of *selja taga* (back+behind) as a complex postposition would take more time. However, the increasing relative frequency of the adverbs may be further explained by viewing the use of individual functions.

Figure 48 depicts the frequencies of *selja taga* (back+behind) when used as an adverb and as a postposition for each complex function as well as the pro-portions of adverbial and postpositional uses within all of the uses as a complex unit.

231
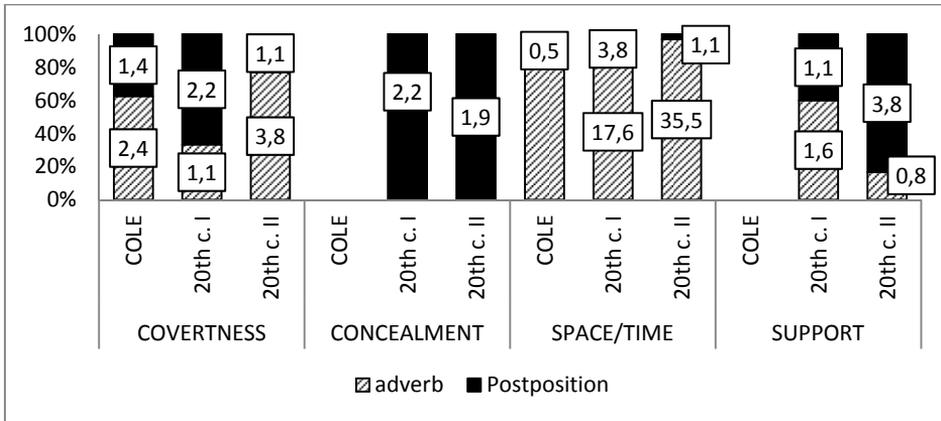
Figure 48. The distribution of the complex postpositions and complex adverbs among the functions of *selja taga* (back+behind) in the observed periods as instances pmw

It can be observed that the distribution of adverbial and postpositional uses is rather different among the individual functions. Most importantly, the spatio-temporal function is used predominantly as a complex adverb and the post-positional use is quite marginal (and even decreased) over all the analyzed time periods. The proportion of postpositional uses remains around 20% in COLE (0.5 instances pmw) and during the 1st part of the 20th century (3.8 instance pmw), but decreases considerably during the 2nd half of the 20th century to 3% (1.1 instances pmw). As was demonstrated in section 4.4, in this function *selja taga* (back+behind) is also predominantly (92%) used as an adverb in contemporary language. As the spatio-temporal function is by far the most common function among the complex units in all the time periods (see Figure 46) and its use increases throughout the 20th century, it is natural that the proportions of the adverbial usage also increase.

The data suggest that the developmental path of the function SUPPORT is more in line with the general principles of grammaticalization. As shown in Figure 46, SUPPORT was not found in COLE. Figure 48 shows that in the 1st half of the 20th century, the function is used as both an adverb (1.6 instances pmw; 60%) and a postposition (1.1 instances pmw; 40%). Based on the available data, it cannot be stated whether this function was first used as an adverb or as a postposition, because both of these parts-of-speech occur only once in the 1900s. However, in the 2nd half of the 20th century, the postpositional use increases to 3.8 instances pmw, which makes up 84% of the uses of this function. Due to the paucity of data, this analysis should however be interpreted with caution. Nevertheless, in the contemporary data SUPPORT is also primarily used as a postposition (80%) (see section 4.4). Thus, in this case, the data suggest a change towards more grammatical uses.

The function COVERTNESS also occurs as both parts-of-speech. However, in this case, there was no clear direction, rather its usage fluctuates over the observed periods. The data suggest that in COLE, COVERTNESS more often

(2.4 instances pmw; 62%) occurred as an adverb. However, in the 1st part of the 20th century, the postpositional use is more common (2.2 instances pmw; 67%), but in the 2nd half of the 20th century, the adverbial use is again more prominent (3.8 instances pmw; 76%). As shown in section 4.4, in contemporary Estonian adverbial and postpositional functions are equally as frequent (54% and 46% respectively).

The function CONCEALMENT, which did not appear in the diachronic data before the 20th century, appears to occur only as a comlex postposition. However, it is also infrequently (8% of the cases where it appears as a complex item) used as a complex adverb. This suggests that this function has developed without an intermediate adverbial phase, and the adverbial uses have developed later.

As for the usages of *selja taga* (back+behind) in other functions, the data suggest that they underwent an adverbial intermediate stage. The very first attestations of *selja taga* (back+behind) as a unit are depicted in Figure 49.
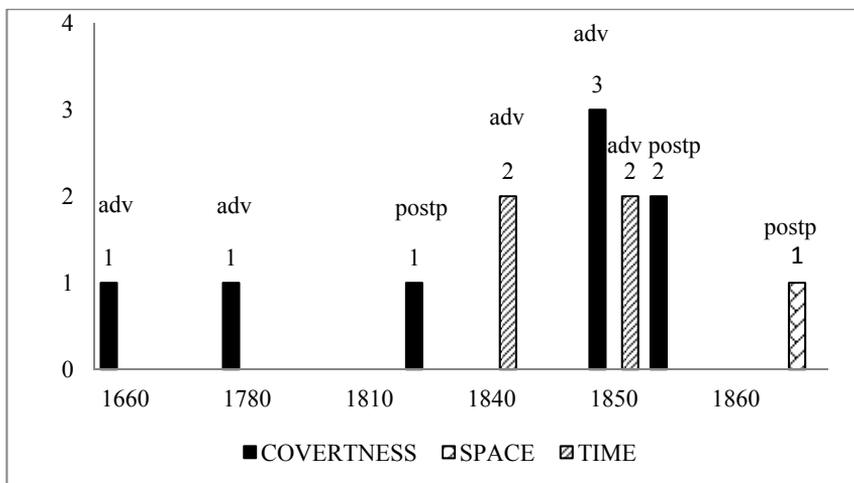
Figure 49. The first attestations of *selja taga* (back+behind) as a complex unit in COLE as absolute frequencies per decade

It must again be acknowledged that the available data is very scarce, therefore the conclusions must be viewed within this context. However, the available data points to the fact that the function COVERTNESS, which seems to be the most clearly established function among the complex units in Old Literary Estonian, appeared first as an adverb and only later as a postposition. When viewing the attestations of COVERTNESS, it can be observed that the first two uses considered as complex units originate from 1662 and 1782 are both used as adverbs (see example (263)). The first complex postpositional usage to express COVERTNESS – which is also the first attested example of *selja taga* (back+behind) as a complex postposition – originates from 1812 (see example (264)). In the 1850s, *selja taga* (back+behind) is used to expresses COVERT-

NESS on 5 occasions (3 instances as a complex adverb and 2 instances as a complex postposition). As postpositional usages do not appear until 150 years later than the first adverbial use, this could mean that the adverbial functions appeared earlier. However, the extremely small numbers of available examples increases the likelihood that this was just a coincidence.

(263) *Hanso   Pertel,   kui   ta   külla-ø   kubja-ø   ammeti-ø*
Hanso   Pärtel   when   s/he   village-GEN   gaffer-GEN   job-GEN
sa-i,   ja   öigus-t   ja   siiwas-d   ello-ø
get-PST.3SG   and   justice-PRT   and   devout-PRT   life-PRT
hakka-s   üllespidda-ma;   siis   wihka-s   te-dda   algamisses
start-PST.3SG   maintain-SUP   then   hate-PST.3SG   s/he-PRT   firstly
nabori-ø   rahwas,   ja   laima-s   te-dda
neighbour-GEN   people   and   blaim-PST.3SG   s/he-PRT
selja-ø   tagga   *kibbedaste.*
back-GEN   behind.LOC   bitterly
'Hanso Pertel, when he got the job as village gaffer and started to uphold justice and devout life then at first the neighbours hated him and slandered him bitterly behind his back.' [COLE, Arwelius, 1782, 73]

(264) *Nee-d   tegge-wad   omma-d   naerowäärt   tempu-sid   ka*
they-PL   do-3PL   own-PL   ridiculous   prank-PL.PRT   also
sejures,   lask-wad   ennes-tel   maks-ta,   naer-wad   te-i-d
thereat   let-3PL   own-PL.ADE   pay-INF   laugh-3PL   you-PL-PRT
*ebbausklikku-d   innimesse-d   **teie-ø**   **selga-ø**   **tagga**,   ja*
superstitious-PL   people-PL   you-GEN   back-GEN   behind.LOC   and
se   haigus   ei   jä-ø   mitte   nenda,   kui   ta   on,
this   disease   NEG   stay-CONNEG   not   still   so as   s/he   be.3SG
waid   ta   saa-b   se-dda   wanne-ma-ks   ja   otse
but   s/he   get-3SG   this-PRT   old-COMP-TRL   and   straight
sepärrast   öäle-ma-ks. (1812-Luce_153)
therefore   vicious-COMP-TRL
'They also make their ridiculous pranks, let people pay them, laugh at you superstitions people behind your back and this disease will not stay the way it is but it becomes older and therefore more vicious.'
[COLE, Luce, 1812, 153]

Figure 49 suggests that the spatio-temporal function does not emerge before 1840s. The complex usages that express temporal and spatial relation (i.e. the spatio-temporal function) have been separated here in order to be able to study the developmental path of the function more scrupulously. It can be observed that in the first available examples of this function, *selja taga* (back+behind) expresses temporal relations only; the locative complex unit appears in the 1860s. It can also be observed that the temporal *selja taga* (back+behind) appears as an adverb only (as in (260) above), a similar adverbial usage is also present in the following decade. The first postpositional usage of this function

appears in the 1860s. This is also the first attestation of the locative *selja taga* (see example (261) above).

As the adverbial and postpositional usages appear temporally very closely together in the data for the spatio-temporal function, it cannot be concluded whether or not the adverbial phase preceded the postpositional usage. However, as stated previously, the postpositional usage is quite rare in the spatio-temporal function in diachronic and synchronic data. Thus, it does not seem very likely that the postpositional usage preceded the adverbial usage. Rather, the postpositional usage seems to have been preferred in instances where *selja taga* (back+behind) expresses spatial or an ordinal function. The available data suggests that *selja taga* (back+behind) used in a temporal sense seems to have preferred to occur as an adverb already in its first attestation. However, the lack of data precludes me from making any further conclusions. What can be stated based on the available data is that it is likely that CONCEALMENT is a more recent function of *selja taga* (back+behind) and probably developed as a complex postposition, without going through an adverbial stage. The data also points to the fact that SUPPORT is increasingly being used as a postposition.

4.8.4.3. Extension of the complex postpositional *selja taga* (back+behind)

In the following, I will discuss the extension of the complex postpositional *selja taga* (back+behind) over the observed period (17th century to 1990s). Following the logic of synchronic analysis, extension is observed in the semantic class of the (pro)nouns (PNs) that precede *selja taga* (back+behind). The PN behaves as a modifier of the body part term in the case of the simple structure and as a complement of the complex postposition in the complex structure.[68] In order to observe the development of each function, where necessary the different functions of *selja taga* (back+behind) are discussed separately. Figure 50 shows the distribution of the semantic classes of the PNs of *selja taga* (back+behind).

---

[68]　It must be noted that in the diachronic data, *selja taga* (back+behind) does not occur with adjectival PNs, i.e. there are no hybrid forms of *selja taga* (back+behind). This is expected, because the hybrid forma are also quite rare (<1%) in the contemporary data (see section 4.2).
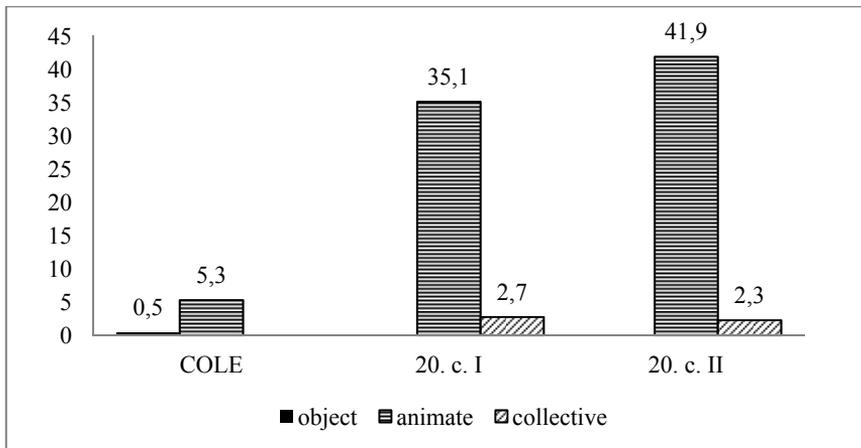
Figure 50. The distribution of the semantic classes of PNs that co-occur with *selja taga* (back+behind) in the observed periods as instances pmw

In contemporary Estonian (see section 4.5.1.3), the PNs of *selja taga* (back+behind) in the diachronic data belong to three semantic classes – animate entities, objects or collectives. The animate PNs are lemmas that refer to either humans or animals. As the body part term *selg* (back) can be used to refer to the body part of humans, animals and to a certain extent objects (as in example (259)), these usages are not considered to represent extension to new contexts. Collective PNs however, are considered as an extension of the use of *selja taga* (back+behind) (as in (265)), because *rahvas* 'nation' is not a noun that can normally modify the body part term *selg* 'back' (see section 4.5.1.3).

(265) *Sultan   kirjuta-s   Prantsuse-ø   kaitselepingu-ø   rahva-ø*
       Sultan   write-3SG   French-GEN   defence agreement-GEN   people-GEN
       *nimel   alla,   aga   <u>rahva-ø</u>   **selja-ø**   **taga**.*
       in the name of   under   but   people-GEN   back-GEN   behind.LOC
       'The sultan signed the French defence agreement in the name of the people but
       behind the people's back.' [AJA1910\ow0004]

The data suggest that collective PNs appear at a later stage. Figure 50 shows that in COLE, *selja taga* (back+behind) occurs either with animate PNs (5.3 instances pmw) or PNs that refer to objects (0.5 instances pmw). However, examples with object PNs do not occur in the 20th century data, but appear again in the contemporary data as a marginal trait (1%). The data suggest that collective PNs of *selja taga* (back+behind) appear in the 1st half of the 20th century (2.7 instances pmw; 7%). The relative frequency of collective PNs remains relatively low (2.3 instances pmw; 5%) during the 2nd half of the 20th century; the number of animate PNs increases along with the number of occurrences of *selja taga* (back+behind) over the same period. As demonstrated in section 4.5.1.3, that refer to collectives make up a modest 9% of the PNs in the con-

temporary data. Thus the data suggest that collective PNs did not appear before the 20th century, and while quite rare, their proportion is slightly larger in the contemporary data. However, this difference might be due to the genre differences between the corpora.

Figure 51 shows that in the diachronic data the collective PN only occurred with complex units (as in example (265)), while the use of animate PNs are possible with both structures. In the contemporary data, it was observed that in rare cases it is also possible to use collective attributes within freely combined phrases (see section 4.5.1).
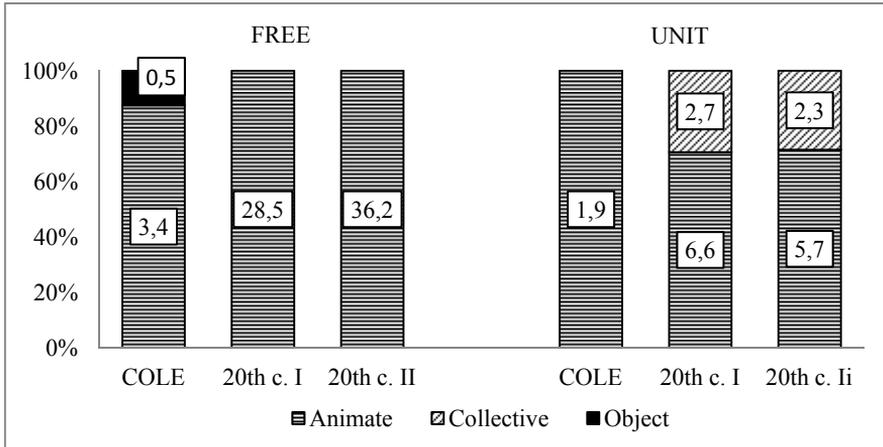
Figure 51. The distribution of the semantic classes of PNs of *selja taga* (back+behind) when used as a freely combined phrase and a complex unit in the observed periods as instances pmw

In case of *selja taga* (back+behind), the distribution of the semantic classes of the PNs among the different functions of *selja taga* (back+behind) is also of interest. When the individual functions are compared (see Figure 52), it can be observed that collective PNs do not co-occur with the function CONCEALMENT, which may be connected to it having developed more recently than other functions of *selja taga* (back+behind). As demonstrated in section 4.5.1, in the contemporary data collective PNs make up 21% of all the PNs in this function. However, despite the fact that the function SUPPORT also only appears in the data from the beginning of 20th century, it is used with both animate and collective PNs. In the case of the other functions, collective PNs appear in the diachronic data during the 1st half of the 20th century. The proportion of collective PNs remains relatively stable for all of the functions. In general in the diachronic data the collective PNs make up a larger amount of complex postpositional uses than in the contemporary data, where collective PNs make up 21% of the instances where *selja taga* (back+behind) behaves as a complex postposition. However, this is probably because the data sample is very small.
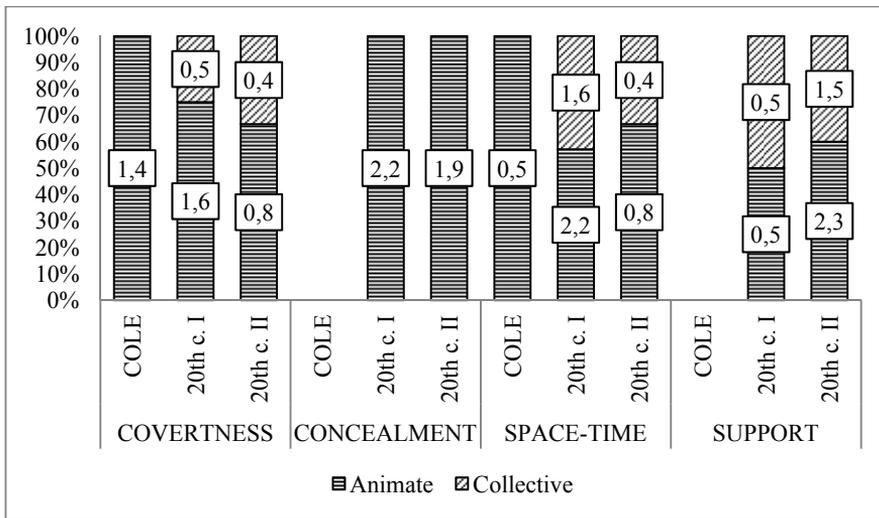
Figure 52. The distribution of animate and collective PNs of *selja taga* (back+behind) in individual functions in the observed periods as instances pmw

Thus, the data suggest that collective PNs of *selja taga* (back+behind) do not appear until the 1st half of the 20th century, and only co-occur with complex units. It is possible that collective PNs just do not turn up in the earlier data but were in use, however it is more likely that they just did not co-occur with *selja taga* (back+behind) before the 20th century. This suggests that *selja taga* (back+behind) as a complex postposition has extended to other contexts over the period investigated. The latter interpretation of the data is also supported by the fact that collective PNs used with *selja taga* (back+behind) are not very common in contemporary language, which may point to the fact that it is a quite recent development.

4.8.4.4. Decategorialization of the complex postpositional *selja taga* (back+behind)

Similarly to the synchronic analysis, decategorialization is observed in the agreement in number between the body part *selg* 'back' and the preceding (pro)noun (PN). As stated in section 2.5.3.2, non-agreement is taken to suggest that the body part phrase is interpreted as a complex unit, and that the preceding nominal is analyzed as the complement of the whole utterance. The distribution of singular and plural PNs is shown in Figure 53.
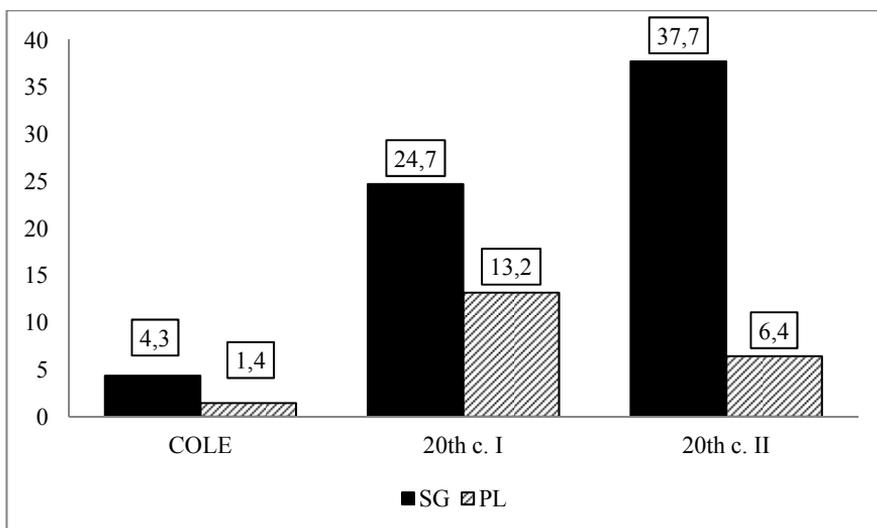
Figure 53. The distribution of singular and plural PNs of *selja taga* (back+behind) in the observed periods as instances pmw

It can be observed in Figure 53 that the plural PNs turn up already in COLE. For instance, in example (264) above) *selja taga* (back+behind) is preceded by plural pronominal *teie* (you.PL.GEN). The data shoes shows that plural PNs are much less frequent than singular PNs in all the corpora. This was expected because the singular, as the unmarked form, is more common than the plural in general, at least when adjacent to a singular body part. However, the data does not suggest an increase in the relative frequency of plural forms over the centuries, but rather a fluctuating dynamic. It can be observed in Figure 53 that in the 1st half of the 20th century, the number of plural PNs increases from 1.4 in COLE to 13.2 instances pmw. However, in the 2nd half of the 20th century, the number of plural PNs drops to 6.4 instances pmw, although the total number of PNs has increased.

In the case of *selja taga* (back+behind) the plural PNs co-occur with both structures, i.e. with complex units and freely combined phrases. For instance, the first occurrence of a plural PN in the data originates from the 1780s and is an example of non-agreement in a freely combined postpositional phrase (see example (266)).

(266) *Taewas   oll-i   selge,   ja   kuu-ø   tärra   tous-is*
      sky   be-PST.3SG clear   and   moon-GEN   shine   rise-PST.3SG
      nende   selja-ø   tagga.
      they-PL.GEN   back-GEN   behind.LOC
      'The sky was clear and the shine of the moon rose behind their back.' [COLE, Arwelius, 1782, 20]

239

In example (266)), *selja taga* (back+behind) is used to express its literal meaning, and is, therefore analyzed as a freely combined phrase. However, the phrase is preceded by the plural pronominal PN *nende* (they.PL.GEN). Such examples were also occurred in the contemporary data where they were explained with the generalization of the meaning of the body part term *selg* 'back' and taken to represent something of an intermediate structure between the freely combined phrases and complex units (see section 4.5.2.3). The diachronic data suggests that the generalization of the body part meaning (at least in this context) was already in use by the end of the 19th century. Such usage may have contributed to the further semantic bleaching of the body part term and blurring of the boundaries of the components of the phrase.

Despite the fact that non-agreement of the body part term and its preceding PN is also found in cases where *selja taga* (back+behind) expresses its literal meaning (BP+LOC), the diachronic data suggest that the plural form is clearly preferred with complex units. Figure 54 shows that while plural PNs are rather marginal in the case of freely combined units, they are much more prominent among complex units in all the observed periods. Amongst complex units, plural PNs make up approximately half of the usages in COLE and during the 1st half of the 20th century (1.0 and 4.4 instances pmw respectively); the proportion of plural PNs decreases in the 2nd half of the 20th century (2.3 pmw; 29%). In contemporary language, plural PNs make up 25% of complex postpositional uses (see section 4.5.2.3). Thus the data suggest that the use of plural forms is decreased among complex units. However, the decrease is probably an effect of differences in the size of the corpora. Figure 54 shows that the occurrence of plural PNs is relatively low among the freely combined units, remaining between 10–30 % in the diachronic corpora. The analysis of the synchronic data suggests that 14% of the PNs are in the plural.
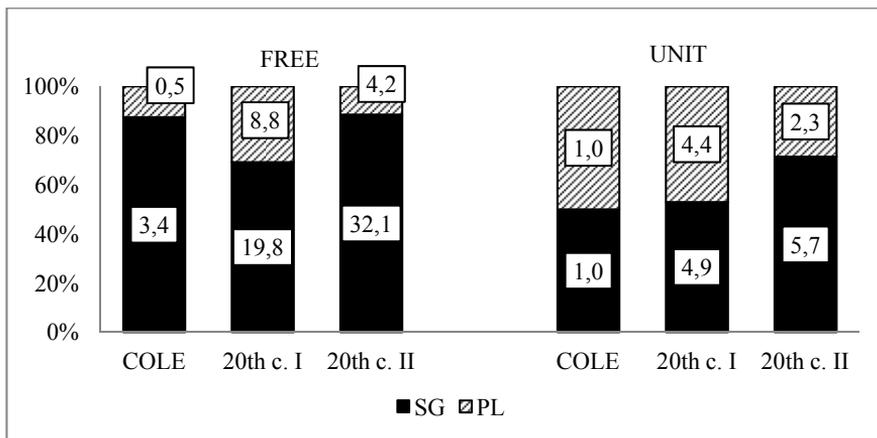
Figure 54. The distribution of singular (SG) and plural (PL) forms of *selja taga* (back+behind) as a freely combined phrase and a complex unit in the observed periods as instances pmw

When agreement is observed within the individual functions, it can be observed that there are some individual differences. Figure 55 shows that only COVERTNESS took plural PNs in COLE. The spatio-temporal function, which is also present in the COLE data, occurs with singular PNs. Nevertheless, it must be taken into account that the spatio-temporal function tends to be realized as an adverb (cf. 4.8.3.2), and therefore PNs are extremely rare in the COLE data for this function (0.5 pmw). The rest of the functions (CONCEALMENT and SUPPORT) are not present in the COLE data. It is remarkable that the function CONCEALMENT, which does not take collective PNs, does take plural PNs already from the first attestations in the 1st half of the 20th century. The data suggest that SUPPORT, which also appears in the data for the 1st half of the 20th century, was not used with plural PNs before the 2nd half of the 20th century. As there is very little data on the individual functions, it is difficult to draw conclusions regarding the dynamic tendencies of the distributions.
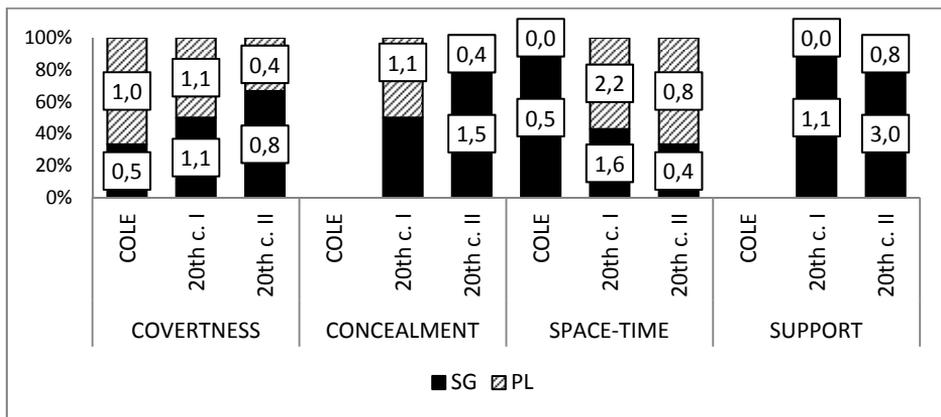
Figure 55. The distribution of singular (SG) and plural (PL) PNs among individual functions of *selja taga* (back+behind) in the observed periods as instances pmw

The fact that the contemporary data suggest that the plural forms are more or less equally represented (20%–27%) among all of the functions (see section 4.5.2.3), should make us even more careful about interpreting the results of the diachronic analysis. The difference between the results of the synchronic and diachronic analysis may suggest that the diachronic analysis does not reflect the development of selja taga (back+behind) realistically. However, based on available data, it was concluded that the plural PNs were preferred with complex units in the diachronic and contemporary data. As in synchronic analysis, this is taken to indicate that the compositional meaning is fading and the phrase is going through decategorialization. However, it seems that the use of plural PNs among complex units (or in general) has not increased; rather, the data suggest fluctuating dynamics over the observed time period.

4.8.4.5. Summary of the diachronic analysis of *selja taga* (back+behind)

The first attestation of *selja taga* (back+behind) in the data originates from 1605 and is analyzed as a freely combined phrase. The data suggest that over the observed periods, the phrase is increasingly used as a complex unit. The total number of instances of *selja taga* (back+behind) in the diachronic data amounts to almost 500 examples, which makes *selja taga* (back+behind) the most frequent of the studied phrases. Despite of that, one must still be cautious when drawing conclusions based on the data, especially when it comes to more detailed analysis (e.g. individual functions), because in this case there are even less examples to analyze, which increases the chance of drawing a false conclusion. Bearing this in mind, the available data suggest the following.

Regarding parts-of-speech, the diachronic data suggest that the complex unit *selja taga* (back+behind) prefers the adverbial function. It has been used mostly (70–80% of its uses as a complex item) as an adverb in all of the analyzed periods. The inclination towards an adverbial use is likely also connected with the fact that the spatio-temporal function, which mostly occurs as an adverb, is the most frequent function in the 20th century and in contemporary language. These results are compatible with the synchronic analysis (see section 4.4).

The data suggest that in the case of *selja taga* (back+behind), the semantic classes of the PNs become more diverse during the 20th century, i.e. collective PNs appear. Similarly to the phrase *käe all* (hand+under), collective PNs co-occur only in such instances of *selja taga* (back+behind) that are analyzed as complex units. Only one function – CONCEALMENT – seems to be of more recent ascent, as exemplified by it taking no collective PNs at all. The diachronic data show that non-agreement appears with both structures – freely combined phrases and complex units. However, non-agreement is more common amongst the complex units in all the observed time periods. This is taken to suggest that the phrases have slowly been losing their compositional meaning, allowing them to decategorialize. These results are compatible with the results of the synchronic analysis.

As for the individual functions, the diachronic data includes no examples of an obvious bridging context between the separate functions. This supports the suggestion made based on the contemporary data (see section 4.3.3) that the individual usage functions root from the same source (the free combination of the body part term and the simple locative postposition), but are not part of the same grammaticalization function, rather they are different branches of development of the source form. The data also suggests that the functions have appeared in different periods of time. COVERTNESS and the spatio-temporal function seem to have been in use earlier than SUPPORT and CONCEALMENT, which appeared at a later stage in the data at the beginning of the 20th century. In the following, the main findings of each function are presented.

To express COVERTNESS is the earliest instance of *selja taga* (back+behind) as a complex unit. It is present already in COLE. The first instance of *selja taga* as a complex postposition also occurs in this function.

The data points to the fact that the development of the complex postpositional *selja taga* expressing COVERTNESS passed through the stage of an adverbial. The spatio-temporal function is also present in COLE but less frequently. However, this function increases considerably throughout the 20th century, and is the most frequent function of *selja taga* (back+behind) as a complex unit in the contemporary data. Despite this, the function is mostly realized as a complex adverb, not a postposition. However, there is some variation within the function, as it also includes a spatial, temporal or ordinal meaning (see section 4.3.3). The available diachronic data supports the assumption that the spatio-temporal function developed as a temporal adverb, and once it was established as a complex unit, it supported the development of the complex locative as well as the positional meaning, which are more likely to be used as postpositions.

Neither SUPPORT nor CONCEALMENT appears in the data before the 20th century. This might suggest they developed more recently. This is supported by the evidence of other aspects of their use I observed. For instance, CONCEALMENT is not used with collective PNs at all, whereas all other functions are. The diachronic data also suggests that the function CONCEALMENT has historically only been used as a postposition. The contemporary data includes some adverbial uses, which make up a marginal amount of examples. This points to the fact that CONCEALMENT has developed as a postposition and did not go through the adverbial intermediate stage, which has been considered to be the developmental path of complex function words that act as adverbs and as postpositions (Habicht, Penjam 2007). However, it is not clear whether *selja taga* (back+behind) expresses SUPPORT first functioned as an adverb or a postposition. The first instances of both uses in the data occur during the same decade, so it is not clear which one appeared first. The data suggest that the proportion of postpositional uses is increasing in this function.

The function OP+LOC analyzed as a free combination of an object part noun and a simple locative postposition is already present in COLE. Despite the fact that the stage where a body part terms serves as an object part is thought to be a vital stage in the development of spatial function words, it does not seem that OP+LOC has contributed to the development of a spatio-temporal function of *selja taga* (back+behind). OP+LOC is not a likely predecessor for the spatio-locative function, because it is extremely scarce in all the studied periods and there are no bridging contexts in either in the synchronic or the diachronic data to link these two functions. For instance, the spatio-temporal function is not used with PNs that refer to object parts.

4.8.5. *Kaela peal* (neck+on)

The phrase *kaela peal* (neck+on) first appears in the corpus in 1601. Similarly to the phrases *käe all* (hand+under) and *selja taga* (back+behind), the first attestation of *kaela peal* (neck+on) appears in the sermons of Müller (267).

(267) *Kuÿ   eb   se-tta   muh-d   keki-t   taha-Ø   teh-a,*
      if   NEG   this-PRT   other-PRT   someone-PRT   want-CONNEG   do-INF
      *se   seis-ka   **tæma-ø   Kaÿla-ø   pæl.***
      this   stand-IMP   s/he-GEN   neck-GEN   on.LOC
      'Lit. If no one else wants to do this it stands on his neck.'
      'If no one else wants to do this it is his responsibility to do it.'
      [COLE, Müller, 1601, s. 13, 28]

In example (267) the phrase *kaela peal* (neck+on) is used to express the notion
of burdening someone with something, and cannot be interpreted literally. This
usage is similar to that in contemporary language, and is considered to be an
example of a complex unit. Thus, once again, the available data does not allow
me observe the very beginning of the developmental path of this phrase. How-
ever, it allows me to gain an insight into the developmental path of the complex
unit of *kaela peal* (neck+on). As the total number of examples in the diachronic
corpora is 66, it must be kept in mind that the available data might not ade-
quately reflect the dynamics of the phrase. In the following, I present an analy-
sis of the diachronic data of *kaela peal* (neck+on), which follows the structure
introduced in section 4.8.1.

4.8.5.1. Frequency of *kaela peal* (neck+on) and the distribution of the simple and the complex structure

The data suggests that phrase *kaela peal* (neck+on) has been highly frequent in
COLE, amounting to a total of 25.6 instances pmw, which makes it the most
frequent of the studied phrases during the period of Old Written Estonian. How-
ever, the data suggest that the use of *kaela peal* (neck+on) dropped dramatically
during the first half of the 20th century and has stayed around 1–2 instances
since (just below 4 instances pmw in the 1st part of the 20th century and above 2
instances pmw in the 2nd half of the 20th century. Although the results of the
diachronic corpora should be compared to the contemporary language data with
caution owing to the vast differences in the size of the corpora (cf. section 3.2),
it should be noted that with 216 occurrences in etTenTen, *kaela peal* (neck+on)
is the least frequent among the studied phrases (see section 4.1). Thus, it seems
that the frequency of *kaela peal* (neck+on) decreased from the 17th century to
the end of 20th century. Decreasing frequency, especially such an abrupt drop as
the data showed, is rather unexpected and not consistent with the general
principles of grammaticalization. Thus, a further investigation of this curious
behavior is needed.
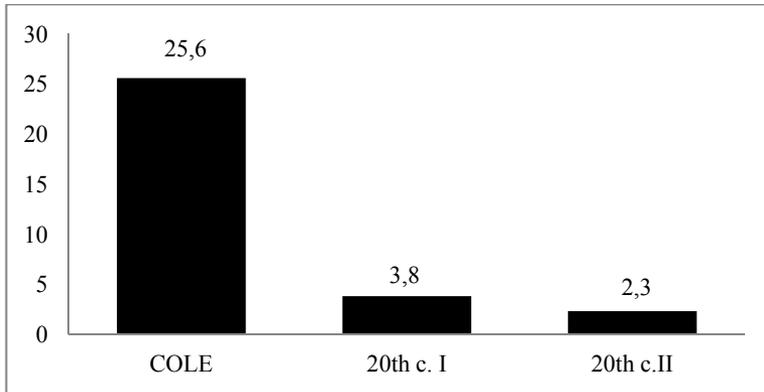
Figure 56. The frequency (pmw) of *kaela peal* (neck+on) in the observed periods

In the following two possible explanations are offered. The first is that the functions of the complex unit *kaela peal* (neck+on) are being overtaken by a competing simple function word. The second explanation is that the frequency of *kaela peal* (neck+on) is over-represented in COLE, because most of the examples originate from the same author. These explanations are not mutually exclusive and most likely both contribute to the explanation for the counter-intuitive finding that the frequency of *kaela peal* (neck+on) decreased over the investigated time period in the data.

The decrease in the frequency of *kaela peal* (neck+on) in the diachronic data is likely associated with the rise of the competing functional word *kaelas* (neck+INE)[69], which seems to be taking over the functions of *kaela peal* (neck+on) as a complex unit (c.f. section 4.3.5). Both of the expressions with the body part term *kael* 'neck' are also analyzable as freely combined structures. Thus, both of them are considered to be relatively new as they are rather transparent, but still clearly established as holistic units. Consider the following examples where *kaelas* (neck+INE) (268) and *kaela peal* (neck+on) (269) both express the notion of being burdened by somebody.

(268) *Kelle-ø*   ***kaela-ø***   *sa*   *nee-d*   *lapse-d*   *taha-d*   *jät-ta?*
       who-GEN   neck-ILL   you   this-PL   kid-PL   want-2SG   leave-INF
       Lit. On whose neck do you want to leave these children?
       'To whose care do you want to leave these children?' [AJA1900\aja0105]

---

[69] Similarly to *kaela peal* (neck+on) and other studied phrases, *kaelas* (neck+INE) also has three forms—lative, locative and separative. In this case, these are formed with the internal locative cases *kaela* (neck+ILL), *kaelas* (neck+INE) and *kaelast* (neck+ELA).

(269) *Aga   noore-l   peremehe-l   ole-ks   inetu   naene   elu-ks*
      but   you-ADE   landlord-ADE   be-COND   ugly   woman   life-TRL
      *aja-ks   **kaela-ø**   **pääl**.*
      time-TRL   neck-GEN   on.LOC
      Lit. But the young landlord would have an ugly woman on his neck for life.
      'But the young landlord would be stuck with an ugly woman for life.'
   [ILU1900\ilu0052]

It is possible that the two expressions, which share an important part of their origin, are competing to express the same notion, and *kaela peal* (neck+on) as the longer, more complex and perhaps more transparent one is losing. Figure 57 shows the overall frequency of *kaelas* (neck+INE) in the observed periods.
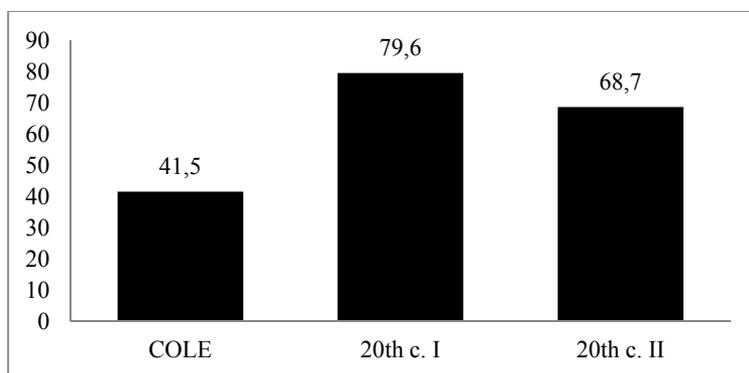
Figure 57. The frequency (pmw) of *kaelas* (neck+INE) in the observed periods

The use of *kaelas* (neck+INE) was already more frequent than that of *kaela peal* (neck+on) in COLE (41.5 and 25.6 instances pmw respectively), and the data also suggest that its use keeps growing during the 20th century (79.6 and 68.7 instances pmw in the 1st and 2nd halves of the 20th century). Thus, it is possible that the decrease in use of *kaela peal* (neck+on) is because of the competition from a shorter and perhaps more grammaticalized alternative.

However, as the literal meaning of the two expressions are not synonymous – *kaelas* (kael+INE) means 'in one's neck' and *kaela peal* 'on one's neck' – the rivalry is best observable among the examples where they are analyzable as holistic units expressing the function BURDEN. The relative frequencies of the phrase *kaela peal* (neck+on) and *kaelas* (neck+INE) as freely combined expressions and holistic units are presented in Figute 58 and Figure 59 respectively.
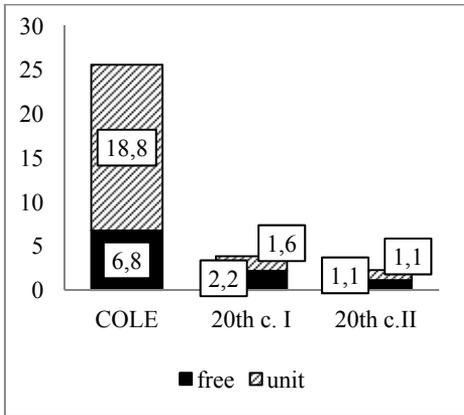
Figure 58. The distribution of the simple and the complex structure of *kaela peal* (neck+on) in the observed periods as instances pmw
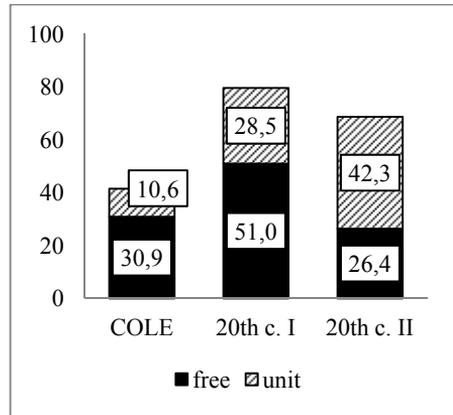
Figure 59. The distribution of the simple and the complex structure of *kaelas* (neck+INE) in the observed periods as instances pmw

The data suggest that it is indeed possible that *kaelas* (neck+INE) is replacing *kaela peal*. Figure 58 shows that in COLE the majority (18.8 instances pmw; 73%) of the examples of *kaela peal* (neck+on) are analyzed as complex units. As the first examples (see example (270)) are already similar to the usages in contemporary language (see example (271)), it seems that *kaela peal* (neck+on) was relatively well established as a complex unit in the 17th century. Moreover, both the relative frequency of the phrase *kaela peal* (neck+on) (25.6 instances) as well as the relative frequency of complex units (18.8) – are more than twice as high as for *käe all* (hand+under) and *selja taga* (back+behind) (see sections 4.8.2. and 4.8.4).

(270) *Sÿn   kule-t   sina   minu-ø A :   Inimene   löhitelt   se*
here   listen-2SG   you   I-GEN   dear   human   shortly   this
eike-ø   Sana-ø   moistus,   sestsinatze-st Laulu-st, Nim
right-GEN word-GEN understanding   this kind-ELA song-ELA nimm
von   vns   lieber H: vnser   Sünd; Se   on   se   eike
von   vns   lieber Herr vnser   Sünd   this be.3SG this right
waña   ninck eßimene   wölgk , me-a   Adam   meddÿ-ø
old   and   first   debt   what-PRT Adam   we-GEN
kaila-ø   pæle   on   tho-nuth.
neck-GEN   on.LAT   be.3SG   bring-PST.PTCP
Lit. Here you my dear man hear briefly the right word of understanding from the song "Nim von vns lieber H: vnser Sünd"; this is the very old and first debt that Adam has brought on our necks.
'Here you my dear man hear briefly the right word of understanding from the song "Nim von vns lieber H: vnser Sünd"; this is the very old and first debt that Adam has brought onto us.' [COLE, Müller, 1604, s. 19, 3]

(271) *Neljandaks,   nn   e-kool   on   tavaõpeta-le*
      fourthly   so-called e-school   be.3SG   regular teacher-ALL
      too-nud   nädala-s   paar   "tasuta lisatundi-ø"
      bring-PST.PTCP   week-INE   couple   free   extra lesson-PRT
      kaela-ø   peale ...
      neck-GEN   on.LAT
      Lit. Fourthly, the so-called e-school has brought a couple of "unpaid extra
      lessons" for a regular teacher on the neck.
      'Fourthly, the so-called e-school has burdened a regular teacher with a couple of
      "unpaid extra lessons".' [www.delfi.ee]

The high frequency of *kaela peal* (neck+on) in COLE may also have been due
to foreign influence. There are no parallel texts in the data that would directly
point to the fact that *kaela peal* (neck+on) is a loan translation, as *käe all*
(hand+under) seems to have been (see section 4.81.). However, the expression
kaela peal (neck+on) can be found in the Estonian-German dictionary of
Wiedemann (1973 [1869]: 178), where it was translated as *Einem auf dem Hals
liegen, zur Last sein* (lit. to lie on one's neck; 'to burden somebody'). As the
German expression has a non-literal and possibly phrasal meaning, it is quite
possible that the meaning of the Estonian *kaela peal* (neck+on) comes from the
German language. However, in this case too (cf. the analysis of *käe all*
(hand+under) in 4.8.2.), it does not seem to be a case of grammatical copying,
but rather a case of a loan translation, because the source is not a grammatical
item but an idiomatic expression. In its earliest periods (16th–18th centuries),
Estonian Literary language was developed by German reverends (Ojutkangas
2001: 107–108), therefore it is possible that the lexicalization of *kaela peal*
(neck+on), which had already clearly taken place by the 17th century, may have
occurred due to or with the help of German influence. In addition to its abuntant
use as a complex unit already in COLE, foreign influence as a trigger of seman-
tic change would also explain the divergence of the simple and complex struc-
ture in contemporary language (see section 4.3.5).

   However, despite of its abundant use as complex item in the 17th to 19th
century, in the 20th century, the phrase almost disappears. At the same time,
kaelas (neck+INE) is used increasingly during the 20th century. Moreover,
Figure 59 indicates that the usages where *kaelas* (neck+INE) is analyzed as a
holistic unit that carries the same meaning as the complex postpositional *kaela
peal* (neck+on) are increasing over the observed periods. In COLE, such exam-
ples appear in 10.6 instances pmw (26%), in the 1st half of the 20th century this
rose to 28.5 instances pmw (36%) and during the 2nd second half of the 20th
century to 42.3 instances pmw (61%). Thus, it is possible that use of the
previously more frequent *kaela peal* (neck+on) ceases because of an increase in
the use of *kaelas* (neck+INE).

   The abruptness of the drop in the frequency of *kaela peal* (neck+on) at the
beginning of the 20th century may point to the fact that the phrase may have
fallen into disfavor of the language planning. The data suggest that a reduction

in the use of *kaela peal* (neck+on) happened rather quickly, which is not suggestive of natural language change, as it is generally considered to be gradual. Of course, we know that the expression is not disappeared from the language completely, because it still occurred in a few examples in the data of the 20th century, and it is also still present in the contemporary data. However, it is difficult to imagine that such a frequently used expression would cease to be used so abruptly. Thus, it may be the case that the language planning favored the use of *kaelas* (neck+INE) over the phrase *kaela peal* (neck+on) to express BURDEN. Although there is no direct evidence about these particular expressions, it is known that Johannes Aavik, an influential language reformer, considered the use of postpositions (including *peal* 'on') a German influence. In the contexts that allow the alternation of cases and postpositions, he suggested to use the former (Aavik 1913: 33; Aavik 1912: 19–20).

The abruptness in the decrease of *kaela peal* (neck+on) over time may be also explained by the nature of the data. As previously suggested, it might be the case that *kaela peal* (neck+on) is over-represented in the COLE data. Indeed, a closer investigation of the data shows that a large proportion of the examples in COLE come from the same author – Georg Müller. This increases the role of a single idiolect and a single type of text (sermons) in the data. However, even if Müller's examples are excluded, the diachronic data still suggest that the overall frequency of *kaela peal* (neck+on) is decreasing. Nevertheless, the decrease is less steep (from 12.6 instances pmw in COLE to 3.8 instances pmw in the 1st half of the 20th century) than if Müller's examples are included. Moreover, in this (smaller) sample, the examples where *kaela peal* (neck+on) behaves as a complex unit still make up just above half of the usages in COLE. Thus, the data still suggests that the phrase must have been a relatively well established complex unit in Old Literary Estonian. Therefore, excluding Müller's examples does not explain the abrupt decrease of *kaela peal* (neck+on) in the 20th century. However, because of the possible methodological issues that may raise from the incompatibility of the corpora (see section 3.2.2), it is also possible that the available data does not realistically depict the frequency and functions of *kaela peal* (neck+on).

4.8.5.2. *Kaela peal* (neck+on) as a complex adverb and a complex postposition

In contemporary Estonian, *kaela peal* (neck+on) as a complex item may serve as an adverb as well as a postposition. Both parts-of-speech also occur in the diachronic data. The distribution of complex adverbs and complex postpositions of *kaela peal* (neck+on) are given in Figure 60.
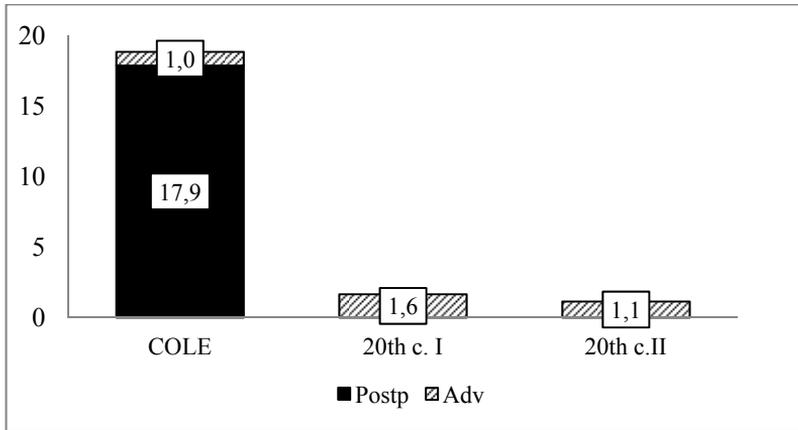
Figure 60. The distribution of the complex postpositions and complex adverbs among the usages of *kaela peal* (neck+on) in the observed periods as instances pmw

Figure 60 shows that in COLE, *kaela peal* (neck+on) is predominately used as a complex postposition (94% of uses (17.9 instances pmw). Unfortunately, the lack of data does not allows me to observe the development of *kaela peal* (neck+on) during the 20th century. The few examples that are present only represent usages as complex adverbs. This may suggest an increase in the adverbial use of *kaela peal* (neck+on) in the 20th century, which is also suggested by the more or less equal distribution of the parts-of-speech in contemporary Estonian. The proportion of adverbial and postpositional usages suggests that adverbial uses make up 41% of all usages as a complex item in contemporary Estonian (see section 4.4). Such a distribution over the observed periods would suggest that postpositional uses preceded adverbial uses of *kaela peal* (neck+on). Figure 61 shows the absolute frequencies of each part-of-speech per decade.
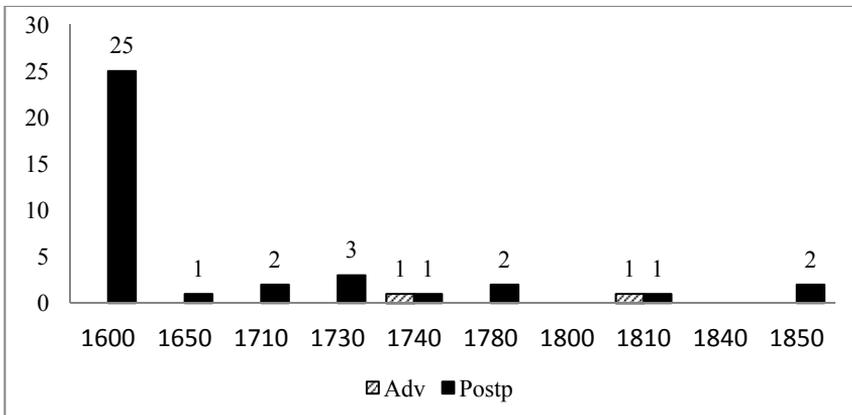
Figure 61. The first attestations of *kaela peal* (neck+on) as a complex unit in COLE as absolute frequencies per decade

250

When the absolute frequencies of the individual examples in COLE are investigated, it can be observed that the majority of examples (25 out of 39) originate from the same decade (1600s). A closer analysis reveals that all of these examples also come from the same author (George Müller). Thus, it is possible that the abundant postpositional use is a trait of Müller's individual speech function. Even with that in mind, it still seems likely that *kaela peal* (neck+on) as a complex item developed first as a postposition, and the adverbial uses were added later. This is first and foremost supported by the fact that adverbial uses are very rare (only two instances out of 39 in COLE). Second, it is indicated by the fact that the very first adverbial use (see example (272)) does not appear before the 1740s, which is 140 years later than the first available examples. Lastly, it is unlikely that *kaela peal* (neck+on) was used as a complex adverb at the beginning of the 17th century and no such uses appeared in Müller's abundant use of the phrase.

(272)   *Öige*   *kül;*   *ütle-wad*   *teise-d,*   *tallopoeg-e-l*   *on*   *raske*
       right   indeed say-3PL   other-PL peasant-PL-ADE be.3PL   hard
       tö   **kaela-ø**   **peäl,**   *millal*   *pea-ks*   *ne-i-l*   *aeg*
       work   neck-GEN on.LOC when   should-COND they-PL-ADE time
       olle-ma   *keik*   *wanna-ø*   *seädusse-ø* *kirja-ø*   *ärralugge-da*
       be-SUP   all   old-GEN   law-GEN   writing-PL   read through-INF
       Lit. True, say others, the peasants have hard work on their necks, when should they have time to read all the old laws.
       'True, say others, the peasants have hard work to do, when should they have time to read all the old laws.' [COLE, Wiis… 1740, 19]

4.8.5.3. Extension of the complex postpositional *kaela peal* (neck+on)

In the following, I will discuss the extension of the complex postpositional *kaela peal* (neck+on). As there is no instances of the postpositional use of *kaela peal* (neck+on) in the 20th century data, the analysis only includes data from COLE. Where appropriate, these results are contrasted with the results of the synchronic analysis. Following the logic of synchronic analysis, extension is observed in the semantic class of the (pro)noun that precedes *kaela peal* (neck+on). However, before I discuss semantic classes of the (pro)nominal component that precedes *kaela peal* (neck+on), I will briefly discuss the instances where the phrase is preceded by an adjective, i.e. the so-called hybrid forms (see section 4.2 and 2.5.3.3).

    The data includes two examples in COLE where *kaela peal* (neck+on) is preceded by an adjective that expresses the characteristics of the body part neck. Such usages provide bridging contexts between the usages as freely combined phrases and complex units. For instance, the use of *kaela peal* (neck+on) in example (273) is semantically similar to the complex units because it is used to express BURDEN, but the adjectival *tugev* ('strong') between *kaela peal*

(neck+on) and the preceding pronoun *temma* (s/he) precludes its analysis as such, as the postposition and the (pro)nominal PN must be immediately adjacent. As mentioned in section 4.2, only one such example could be found in the contemporary data.

(273) *Selle-ø* **_tugeva-ø_** **_kaela-ø_** **_peale_** *pane-b* *talu* *kogu*
this-GEN strong-GEN neck-GEN on.LAT put-3SG farm all
oma-ø *au-ø,* *oma-ø* *kuulsuse-ø,* *terve* *küla-ø,*
own-GEN honour-GEN own-GEN glory-GEN whole village-GEN
kogu *valla-ø* *ees.*
whole parish-GEN in front
'On this strong neck the farm places all its honour, its glory, in front of the whole village whole parish.' [ILU1900\ilu0020]

Regarding the semantic classes of the PNs, the COLE data suggest that when used as a complex postposition, *kaela peal* (neck+on) clearly prefers animate PNs. Animate attributes occur in 95% of the usages (21.2 instances pmw). However in two examples (1.0 instances pmw), the PN refers to a collective. Thus, the collective PNs that are infrequent (10%) are still possible in contemporary language (see section 4.5.1.1). The first such example originates from the year 1605 (see example (274)), with the phrase preceded by the PN *Inimeße Suggu* 'mankind', which refers not to a single person but a collective or even an abstract notion. Both of the examples that include a collective PN represent the complex structure. Thus, it seems that the collective PN is only associated with complex units, which is in line with the results of the contemporary analysis. As a free unit, *kaela peal* (neck+on) also occurs with PNs that refer to object part (see example (275)), which are also present in the contemporary data as a marginal usage function (3%).

(274) *Sÿßkit* *eb* *olle-ø* *næma-t* *se-dda* *mitte*
nevertheless NEG be-CONNEG they-PL this-PRT not
moist-nuth, *kust* *doch* *keick* *hedda* *ninck* *willetzus*
understand-PST.PTCP where yet all ill and destitution
se *Waÿse-ø* **_Inimeße-ø Suggu ø_** **_Kaila-ø_** **_pæle_** *tulle-b, …*
this poor-GEN human-GEN kind-GEN neck-GEN on.LAT come-3SG
Lit. Nevertheless they have not understood from where all the ill and destitution comes onto the neck of the poor human kind.
'Nevertheless they have not understood from where all the ill and destitution comes on the poor human kind.' [COLE, Müller, 1605, s. 24 5]

(275) *Küünita-si-n* *lambi-ni* *ja* *pan-i-n* *vati-ø* *selle-ø*
crane-PST-1SG lamp-TER and put-PST-1SG cotton-GEN this-GEN
kaela-ø **_peale_**.
neck-GEN on.LAT
'I reached the lamp and put the cotton onto its neck.' [ILU1990\ilu0523]

The collective PN was already present in COLE, which is not the case for any of the other studied phrases. This supports the claim made above that *kaela peal* (neck+on) must have been a rather established complex unit at that time. Of course, the more varied uses of PNs in the case of *kaela peal* (neck+on) may be due to the fact that there are more data regarding this phrase, but it is likely that the higher frequency of the phrase itself is a consequence of it being more grammaticalized than the other studied phrases in Old Literary Estonian.

4.8.5.4. Decategorialization of the complex postpositional
kaela peal (neck+on)

In this section, I discuss the decategorialization of the complex postpositional *kaela peal* (neck+on), which is observed in the agreement in number between the body part *kael* 'neck' and the preceding (pro)noun (PN). As was stated above (cf. 2.5.3.2), the non-agreement are taken to suggest that the body part phrase is interpreted as complex unit, and that the preceding nominal is analyzed as the complement of the whole utterance. Similarly to extension, decategorialization (see previous section) may be observed in COLE data only.

The data suggests that the plural PNs are present in COLE, where they make up 41% of the data (9.2 instances pmw out of 22.2) of complex postpositions. As was mentioned, the data does not allow me to observe the distributions of the number of PNs during the 20th century. However, it was observed that both forms are present in the contemporary data, where the proportion of plural forms is 31%. Thus, the data suggests that the decategorialization was even more wide spread in Old Literary Estonian.

A closer analysis of the examples that feature non-agreement reveals that the plural forms occur with both structures. However, the data suggests that the plural forms are clearly more likely to occur with complex units – about half of the usages as complex postpositions (8.7 instances pmw) are used with a plural PN. Such usages already occur in the earliest available examples, e.g. example (270) from 1604. In this example we see that plural PN (*meie* 'we') precedes the body part term, which is in the singular form (*kaela*), resulting in non-agreement between the PN and its supposed head. In such cases the pronoun *meie* 'we' stands form the human kind who is the bearer of debt that Adam has brought upon. Such usages indicate reanalysis of the phrasal structure, so that the preceding pronoun no longer modifies just the body part term but the whole utterance *kaela peal* (neck+on) which is analyzed as a holistic unit that carries a distinct meaning.

Nevertheless, in the diachronic data, there is a single example were plural PN is used with an instance of *kaela peal* (neck+on) which is used in its literal meaning and, therefore, analyzed as freely combined unit (see example (276). Such usages also occurred with *selja taga* (back+behind) (see 4.8.4.4) and *külje all* (side+under) (4.8.3.4). In this instance, the example illustrates a usage where

the body part term is used to refer to (animal's) neck in general (cf. 2.5.3.2), and *kaela peal* (neck+on) is not lexicalized. Therefore, the non-agreement is not taken to reflect reanalysis of the phrasal structure. As was also discussed above (in section 4.5.2.5), in contemporary language, there is also just one ambiguous example that represents non-agreement among freely combined phrases.

(276) *Seal   pan-d-i   ratsuse-ø   otsa-d   **oos-te**   **kaela-ø***
      there   put-IMPS-PST   bridle-GEN   end-PL   horse-PL.GEN   neck-GEN
      peale   *ja*   *ae-t-i*   *ne-i-d*   *mine-ma.*
      on.LAT   and   chase-IMPS-PST   they-PL-PRT   go-SUP
      'There the ends of the bridle were put on the neck of the horses and they were chased away.'
      [COLE, Kreutzwald, 1850, 1833]

Thus, the diachronic data suggest that the non-agreement of the body part term *kael* 'neck' and its PNs was already quite common in Old Literary Estonian. It is likely that this is in part connected with the genre of the texts in COLE, because *kaela peal* (neck+on) often co-occur with the pronoun *meie* 'we', which stands for humankind. Humankind as a target of various burdens is characteristic of religious discourse. In both Old Literary Estonian and contemporary language, it is also possible that non-agreement occurs in the simple structure, but such usages are rare in the data.

4.8.5.5. Summary of the diachronic analysis of *kaela peal* (neck+on)

The first attestation of *kaela peal* (neck+on) originates from 1601 and represents a use as a complex unit. The data does not suggest that its use as a complex item has increased over time, but rather that the use of *kaela peal* (neck+on) disappeared almost entirely during the 20th century. The absolute number of all usages in the diachronic corpora is only 66, therefore the analysis should be interpreted with caution. As the phrase is still present in contemporary language (in etTenTen), it is obvious that *kaela peal* (neck+on) has not fallen completely out of use. However, it should be noted that the phrase *kaela peal* (neck+on) is the least frequently used among all the studied other phrases in contemporary data.

    The diachronic analysis suggests that in comparison with the other studied phrases, *kaela peal* (neck+on) was used rather frequently in Old Literary Estonian. Nevertheless, it is likely that the phrase is over-represented in the COLE data, because a large proportion of examples originate from the same author. Even if those examples are excluded from the data, the remaining data suggest that *kaela peal* (neck+on) was quite well established as a complex unit in Old Literary Estonian, but for some reason there are only a few examples of the phrase in the 20th century data, which was rather unexpected. *Kaela peal* (neck+on) might have been replaced by the competing simple function word

kaelas (neck-INE), which when analyzed holistically, can be used to express a meaning to that of *kaela peal* (neck+on).

The diachronic data suggests that as a complex unit, *kaela peal* (neck+on) was first used as a postposition. This is indicated by the abundant use as complex postpositions in the diachronic data, and the fact that the postpositional uses seem to be earlier. This is also supported by the fact that the adverbial use increases in the more recent data. As a complex postposition, *kaela peal* (neck+on) stands out as a unit that was already rather frequently complemented by a plural PN in Old Literary Estonian. Also, the phrase was already used with a collective PN in COLE, which is unprecedented among the other studied phrases.

4.8.6. *Käe kõrval* (hand+beside)

The phrase *käe kõrval* (hand+beside) first appeared in the corpus in 1818 in "Pühhapäwa Wahhe-luggemissed" ('Sunday Readings'), by Otto Willem Masing. This first attestation of *käe kõrval* (hand+beside) is a lexicalized usage (see example (277)), as are the most of the examples in the diachronic data.

(277)   *On*   *siis*   *ka*   *mitto*   *kõrda-Ø*   *se-dda*   *nähh-a,*   *et*
     be.3SG   then   also   several   time-PRT   this-PRT   see-INF   that
     nõuka-l   *perremehhe-l*   *wiis, kuus, seitse*   *sadda,*   *ehk*   *peäle*
     clever-ADE   landlord-ADE   five six   seven   hundred   maybe   on
     tuhhat-ki   *põtra-ø,*   *ke-dda*   *keik*   *kui*   *omma-ø*   *koddo-ø*
     thousand-CL   moose-PRT   who-PRT   all   as   own-GEN   home-GEN
     lojus-sid   *tunne-b;*   *igga*   *ühhe-le*   *temma-ø*   *nimme-ø*
     animal-PL.PRT   know-3SG   every   one-ALL   s/he-GEN   name-GEN
     anna-b,   *ja*   *te-dda*   *nimme-ø*   *kaupa*   *omma-ø*   **käe-ø**
     give-3SG   and   s/he-PRT   name-GEN   by   own-GEN   hand-GEN
     kõrwa   *kutsu-b.*
     beside.LAT   call-3SG
     Lit. It has been seen many times that a clever landlord has five, six, seven hundred or even over a thousand moose, all of whom he knows as his own domestic animals, gives each a name and calls it by its name to the side of his hand.
   'It has been seen many times that a clever landlord has five, six, seven hundred or even over a thousand moose, all of whom he knows as his own domestic animals, gives each a name and calls to his side by its name.' [COLE, Masing, 1818, 16]

The total number of examples of *käe kõrval* (hand+beside) in the diachronic data is only 25. Thus, the following analysis will be rather qualitative than quantitative. In both cases, the paucity of examples means that all conclusions must be viewed with caution. In the following, I present the analysis of the dia-

chronic data of *käe kõrval* (hand+beside), which follows the structure introduced in section 4.8.1.

4.8.6.1. Frequency of *käe kõrval* (hand+beside) and the distribution of the simple and the complex structure

Although examples were few, it seems that the use of *käe kõrval* (hand+beside) is increasing over the observed periods – COLE includes 1.4 instances (pmw) of *käe kõrval* (hand+beside) but the relative frequency of the phrase had risen to 2.7 instances pmw in the 1st half of the 20th century and 6.4 instances pmw in the 2nd half of the 20th century (see Figure 62).
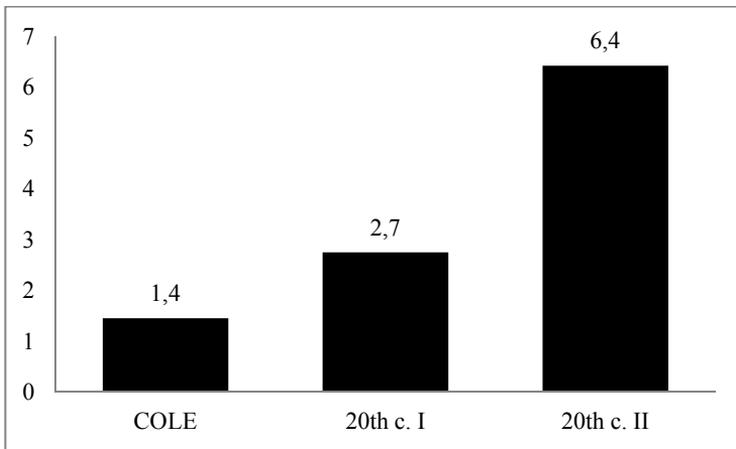
Figure 62. The frequency (pmw) of *käe kõrval* (hand+beside) in the observed periods

When the simple and the complex structure are analyzed separately, the results during the period of Old Literary Estonian were similar to that of the contemporary analysis (see section 4.3.4), i.e. *käe kõrval* (hand+beside) was predominantly used as a complex unit. However, in the contemporary data *käe kõrval* (hand+beside) was also used to express the functions BESIDE and ACCOMPANIMENT (see section 3.4.3), whereas the diachronic corpora mostly includes examples where *käe kõrval* (hand+beside) expresses the function BESIDE (see example (278)). However, there are also a few examples in the diachronic data from the 2nd half of the 20th century that express ACCOMPANIMENT (see example (279)). Although it could be expected that the more abstract ACCOMPANIMENT is a more recent development, there is not enough data available to be able to conclude this with certainty.

(278) *Siis   sammu-b   ta   Hermanni-ø   **käe-ø**   **kõrval***
then   step-3SG   s/he   Hermann-GEN   hand-GEN   beside.LOC
kodu-ø   poole.
home-GEN   toward
Lit. Then s/he walks beside Hermann's hand towards home.
'Then s/he walks beside Hermann towards home.' [ILU1910\ilu0020]

(279) *Võta-me   te-i-d   **käekõrvale**,   vii-me   Euroopa-sse*.
take-1PL   you-PL-PRT   hand.beside.LAT   bring-1PL   Europe-ILL
Lit. We will take you beside hand, take you to Europe.
'We will take you by our side, take you to Europe.' [AJAE1990\tat0661]

The freely combined phrases make up less than 1% of examples in the contemporary data; in the diachronic data, there is only a single example (0.4 instances pmw) of such a usage, which appears in the 2nd half of the 20th century. Thus, the data suggest that *käe kõrval* (hand+beside) already functioned as a complex unit in Old Literary Estonian. Of course, it must have been possible to combine the components of the phrase to form a freely combined postpositional phrase back then. However, it is likely that no such examples occur in the corpus because of the extremely small number of examples. Moreover, similarly to the phrase *käe all* (hand+under) (see section 4.8.2), *käe kõrval* (hand+beside) as a freely combined phrase denotes quite a specific meaning (see example (280)). Therefore, the rarity of such examples is expected.

(280) *Läbi   köögiakna-ø   paist-is   länguvajunud*
through   kitchen window-GEN   appear-PST.3SG   slumped
puukuur,   selle-ø   ees   toksi-s   perenaise-ø
woodshed   this-GEN   infront   knock-PST.3SG   housewife-GEN
vend,   sõja-s   jala-ø   kaota-nud   keskealine   mees
brother   war-INE   leg-GEN   lose-PST.PTCP   middle-aged   man
pliidipu-i-d,   kirve-ga   alati   nii   täpselt   puu-d
stove wood-PL-PRT   ax-COM   always   so   accurately   wood-PRT
*hoid-va   **käe-ø**   **kõrvale**   taba-des,   et   Hendriku-l*
hold-PTCP.GEN   hand-GEN   beside.LAT   hit-GER   that   Hendrik-ADE
alati   tekki-s   tahtmine   silm-i   sulge-da –   nüüd
always   come up-PST.3SG   will   eye-PL.PRT   close-INF   now
löö-b   sõrme-de-le.
hit-3SG   finger-PL-ALL
'Through the kitchen window one could see a slumped woodshed, in front of it the housewife's brother, a middle-aged man who had lost his leg in the war, was hacking wood, always hitting with the axe so accurately beside the hand that held the log that Hendrik always wanted to close his eyes – now he will hit his fingers.' [ILU1970\ilu0072]

4.8.6.2. *Käe kõrval* (hand+beside) as a complex adverb and a complex postposition

In contemporary Estonian *käe kõrval* (hand+beside) is mostly used as a complex adverb. However, complex postpositional uses are also possible but infrequently used (see section 4.4). As there is very few examples in the diachronic corpora, the distribution of the adverbial and postpositional uses of *käe kõrval* (hand+beside) are viewed in absolute numbers (Figure 63).
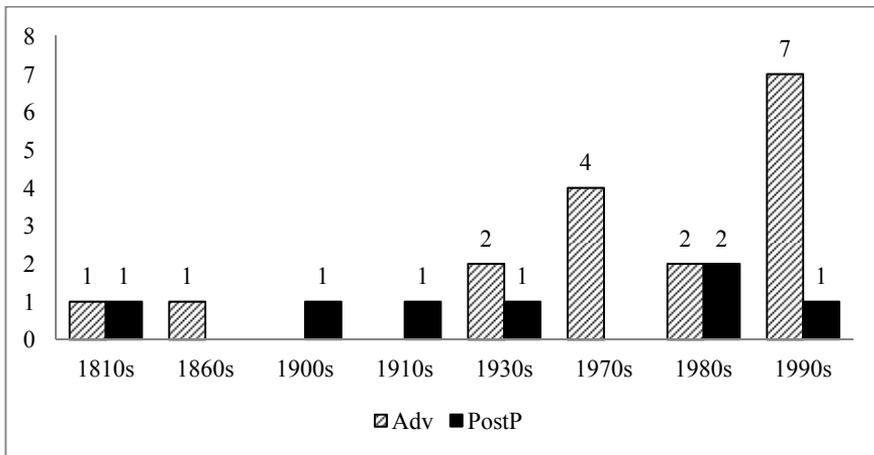
Figure 63. The first attestations of *käe kõrval* (hand+beside) as a complex unit in the diachronic data in absolute frequencies per decade

There are too few examples to conclude whether *käe all* (hand+beside) tends to be realized as an adverb or as a postposition during the 19th century and the 1st half of the 20th century. However, the data suggests that *käe kõrval* (hand+beside) prefers the adverbial function in the 2nd half of the 20th century – in the 1970s all of the examples of complex units are adverbs, in the 1980s both adverbial and postpositional uses are equally represented, but in the 1990s the adverbial use is again more common (7 instances vs. 1 instance). The synchronic analysis yielded similar results (see section 4.4). Although *käe kõrval* (hand+beside) is predominantly used as a complex adverb in the more recent periods, it could be hypothesized that the complex adverbial uses preceded the complex postpositional uses. However, the data does not offer a clear answer. Figure 63 shows that adverbial and postpositional uses appear at the same time. The first complex postpositional use was exemplified in example (277), the first example of adverbial use appeared in the same year in the same text by the same author (Masing, 1818; see example (281)).

(281) *Se-st   hirmsa-st   ja   kardetawa-st   ellaja-st   Wanna-s*
      this-ELA   scary-ELA   and   feared-ELA   beast-ELA   Old-INE
      Testamenti-s   Jobi-ø   ramatu-s,   40   peatükki-ø   25   salmi-st
      Testament-INE   Job-GEN   book-INE   40th chapter-GEN   25th pslam-ELA
      kunni   41   peatükki   wimase-st   salmi-st   sadik,   paljo-gi
      to   41st chapter-GEN   last-ELA   psalm-ELA   since   many-CL
      õppe-ta-kse,   mis   nen-de,   pärrast,   ke-l
      teach-IMPS-PRS   what   they-PL.GEN   because   who-ADE
      *Piibliramatu-d   **käe-ø**   **kõrwas**   ei   pea-ks*
      Bible book-PRT   hand-GEN   beside.LOC   NEG   should-COND
      olle-ma,   tükki-ø   kaupa   seie   tahha-me   jätka-da.
      be-SUP   piece-GEN   by   here   want-1PL   continue-INF
      Lit. Much is taught of this terrible and feared beast in the Old Testament in the
      book of Job from the 25th psalm of the 40th chapter to the last psalm of the 41st
      chapter that for those who do not happen to have the Bible beside hand we will
      here continue chapter by chapter.'
      'Much is taught of this terrible and feared beast in the Old Testament in the book
      of Job from the 25th psalm of the 40th chapter to the last psalm of the 41st chapter
      that for those who do not happen to have the Bible at hand we will here continue
      chapter by chapter.' [COLE, Masing, 1818, 102]

Thus, although there is extremely little data, it seems that both parts-of-speech
were already used by the first attestations of *käe kõrval* (hand+beside). How-
ever the order of the development of the adverbial and postpositional functions
remains inconclusive.

4.8.6.3. Extension and decategorialization
of the complex postpositional *käe kõrval* (hand+beside)

In this section, I discuss the extension and decategorialization of *käe kõrval*
(hand+beside). Both of them are observed in the features of the (pro)nominals
of *käe kõrval* (hand+beside). As in previous sections, extension is observed via
the semantic classes of the PN and decategorialization in the non-agreement in
number between the body part term *käsi* 'hand' and the PN. The following
analysis will demonstrate that neither of these notions are extensively estab-
lished in the diachronic data of *käe kõrval* (hand+beside).[70]

As discussed in section 4.5.1.4), the PNs of *käe kõrval* (hand+beside) are not
particularly diverse in contemporary language. The same applies for earlier
examples. The contemporary data suggest that *käe kõrval* (hand+beside) is pre-
dominantly used with human PNs, though in rare cases the PNs referred to a
collective or an abstract notion (see section 4.5.1.4). In the diachronic data, only

---

[70] It must be noted that in the diachronic data, *käe kõrval* (hand+beside) does not occur
with adjectival modifiers at all. A few (2 out of 780) instances of such usages can be
however found in contemporary data (see section 4.2).

human PNs occur. As both of the contemporary examples of inanimate PNs of *käe kõrval* (hand+beside) occurred with complex items, it seems that the complex postpositional *käe kõrval* (hand+beside) has extended to new contexts. However, because of the extreme marginality of such examples and the lack of historical data, the possibility remains that the occurrence of these examples might be trivial to the development of *käe kõrval* (hand+beside).

The diachronic data do not include any instances of non-agreement between the body part term and the PN. As demonstrated above, non-agreement is also rather infrequent (12%) in contemporary data. As discussed in section 4.5.2.4, in contemporary data plural PNs only co-occur with instances of *käe kõrval* (hand+beside) analyzed as complex units (or hybrid forms). Thus, it may be concluded that the appearance of plural PNs suggests decategorialization. However, due to the lack of data, the possibility remains that such usages were also present historically, but did not occur in the corpus.

4.8.6.4 Summary of the diachronic analysis of *käe kõrval* (hand+beside)

The first attestation of *käe kõrval* (hand+beside) originates from 1818 and represents use as a complex unit. The data suggest that by the first attestations in the 19th century, *käe kõrval* (hand+beside) was already conventionalized as a complex unit. In the data of the 19th and 20th centuries, there is only one instance where *käe kõrval* (hand+beside) is used as a freely combined unit. It is possible that this is because of the specific meaning of the freely combined phrase, as well as the general infrequency of *käe kõrval* (hand+beside) in the diachronic data. As the total number of available examples is only 25, the following conclusions should be viewed with caution.

The available data suggest that the proportion of complex units has not increased over the observed periods, because the phrase is mostly used as a complex unit. However, contrasting this findings with that of the contemporary data, it can be observed that the amount of examples where *käe kõrval* (hand+beside) was used to express the notion of ACCOMPANIMENT has increased. As ACCOMPANIMENT is a more abstract notion than BESIDE, the data suggest further grammaticalization. However, the extremely small data sample may have increased the role of coincidence. Thus, it may be the case that such usages were more frequent in the 19th and 20th century, but are not fully represented in the corpus because the function is in general is quite rare.

Regarding the part-of-speech, *käe kõrval* (hand+beside) is mostly used as a complex adverb in contemporary language (see section 4.4). The diachronic data suggest that this is also the case in the 2nd half of the 20th century. Unfortunately, the data does not allow me to draw conclusions on the chronological order of the adverbial and postpositional stages in the developmental path of *käe kõrval* (hand+beside). Both parts-of-speech appear in the data at the same time (1818) in the same text by the same author – Otto Willem Masing.

As there is very little data and because *käe kõrval* (hand+beside) is mostly used as a complex adverb, it is very difficult to draw conclusions about *käe kõrval* (hand+beside) as a complex postposition. However, the available data does show that *käe kõrval* (hand+beside) is historically only used with human PNs and plural PNs. Inanimate and plural PNs are however present in contemporary data to a certain extent. Thus, it seems that the immediate contexts of *käe kõrval* (hand+beside) may have expanded. However, the absence of such features in the earlier periods may be also due to the small number of examples and or differences in the genres. As the contemporary data comes from the internet and thus is often only partly edited, it is possible that it contains more innovative language use.

4.8.7. *Jalge all* (feet+under)

The first attestation of *jalge all* (feet+under) also appears in 1818 in "Pühhapäwa Wahhe-luggemissed" ('Sunday Readings') by Otto Willem Masing. In this example *jalge all* (feet+under) represents the simple structure (see example (282)). However, the phrase is used with the verb *tallama* 'tread', which is a frequent collocate of the complex unit in contemporary language (see section 4.6.6.2). Thus, this example serves as a bridging context between the functions BP+LOC and OPPRESSION.

(282) *Kui temma mahha heit-nud, ning kegi te-dda siis weel*
     if   s/he   lie down-PST.PTCP   and   someone   s/he-PRT   then   more
    hakka-b sundi-ma, et ta wäggiselt ja ülle joudu-ø
    start-3SG   force-SUP   that   s/he   by force   and   over   weight-PRT
    omma-ø koorma-ga emale pea-b minne-ma: siis temma
    own-GEN   load-COM   away   must-3SG   go-SUP   then   s/he
    tulle-b perro-ks, ja nenda hullu-ks, et ta ülles
    come-3SG   skittish-TRL   and   so   mad-TRL   that   s/he   up
    *tõuse-b, omma-ø sundija-d kohhe **jalg-e alla***
    stand-3SG   own-GEN   forcer-PRT   just   foot-PL.GEN   under.LAT
    talla-b, ja te-dda purruks sõtku-b.
    tread-3SG   and   s/he-PRT   into pieces   tread-3SG
    'When it has lain down and someone then still starts to force it that it by force
    and beyond its strength must go away with its load then it becomes skittish and
    so mad that it stands up, instantly tramples its forcer under its feet and stomps
    him into pieces.' [COLE, Masing, 1818, 9]

Before the first instances of *jalge all* (feet +under), the combination of *jalad* (feet) and *all* (under) was formed with the longer agglutinative genitival form *jalga-de all* (foot+GEN + under) (see also section 4.1). Such usages were already present in Müller's sermons of the 1600s. Since phrases formed with the longer genitive are outside of the scope of this study, these will not be discussed

any further. However, it should be mentioned that the data suggest that although the shorter form (*jalge all*) appears later, its use increases over the studied periods and by the 2^(nd) half of the 20^(th) century, it has capped the longer forms (*jalgde all*). The distributions of the phrases formed with the short (fusional) genitive and long (*de-*) genitive are depicted in Figure 64.
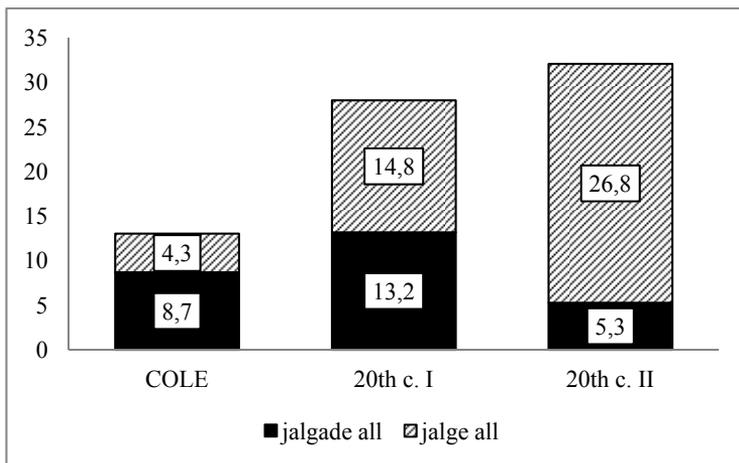

Figure 64. The distribution of *jalgade all* and *jalge all* (feet+under) in the observed periods, relative frequency

As *jalge all* (feet+under) only appears in the beginning of 19^(th) century, it is less frequent (4.3 instances pmw) in COLE than the longer *jalgade all* (8.7 instances pmw). However, the two forms are used more or less equally in the 1^(st) half of the 20^(th) century (13.2 and 14.8 instances pmw). By the 2^(nd) half of the 20^(th) century, the use of *jalge all* (feet+under) has become dominant (26.8 instances pmw), making up 83% of uses. The increasing frequency of *jalge all* suggests that it has more potential to grammaticalize.

The total number of examples of *jalge all* (feet+under) in the diachronic data is 108. Thus, it is the second most frequent of the studied phrases next to *selja taga* (back+behind). Nevertheless, considering that these examples cover many centuries of development, the following analysis should still be interpreted with caution. In the following, I present the diachronic analysis of *jalge all* (feet+under), which follows the structure presented in 4.8.1.

4.8.7.1. Frequency of *jalge all* (feet+under) and the distribution of the simple and the complex structure

The frequency of *jalge all* (feet+under) was already presented in Figure 64. In this section, I discuss the distribution of the simple and the complex structure in the observed periods (see Figure 65).
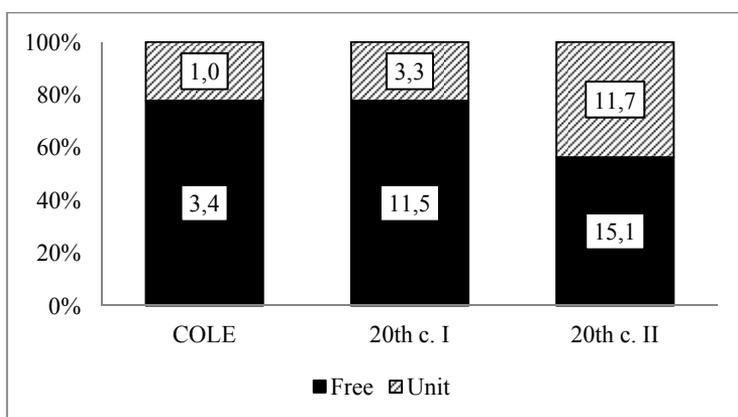
Figure 65. The distribution of *jalge all* (feet+under) as a freely combined phrase and a complex unit, instances pmw

The data show that both of the structures – the freely combined phrase and the complex unit – are present in all the studied periods, and that the proportion of complex units increases over time. In COLE and the 1st half of the 20th century, the amount of instances that are analyzed as complex units is approximately 20% (1.0 out of 3.4 instances pmw). In absolute numbers, there were just two examples of lexicalized uses of *jalge all* (feet+under) in COLE and 6 in the data from the half of the 20th century. However, in the half of the 20th century the complex units make up more than 40% (11 instances pmw out of 26.8) of all the usages of *jalge all* (feet+under). However, freely combined units (for an example see (283)) are more frequent than complex units in all the observed periods. Thus, it can be concluded that although the frequency of the complex units increases over time, the instances where *jalge all* (feet+under) is analyzed as a freely combined unit remain more prominent.

(283) *Metsa-d   kohha-si-d   ja   ohka-si-d   wasto   suretükki-ø*
       forest-PL   rustle-PST-3PL   and   sigh-PST-3PL   on   cannon-GEN
       mürrina-st,   innimes-te   häddakarrimise-st   ja   hoos-te
       grumble-ELA   people-PL.GEN   screaming-ELA   and   horse-PL.GEN
       *hirnumise-st;   maa   olle-ks   **jalg-e**   **al***
       horselaugh-ELA   land   be-COND   foot-PL.GEN   under.LOC
       otsekui   liku-nud   se-st   hirmsa-st   põrrutamise-st.
       as   move-PST.PTCP   this-ELA   terrible-ELA   shake-ELA
       'The woods rustled and sighed back the roar of cannons, screaming of people
       and neighing of horses; it felt as if the ground moved under feet from this terrible
       shaking.' [COLE, Suve Jaan, 1841, 32]

As discussed in section 4.3.6, in contemporary language *jalge all* (feet+under) occurs in two functions that do not share a common developmental path in more ways than the same source form. Thus, where possible, these functions will also

be discussed separately in the diachronic analysis. In the following, the distribution of the functions in the observed periods is discussed (see Figure 66).
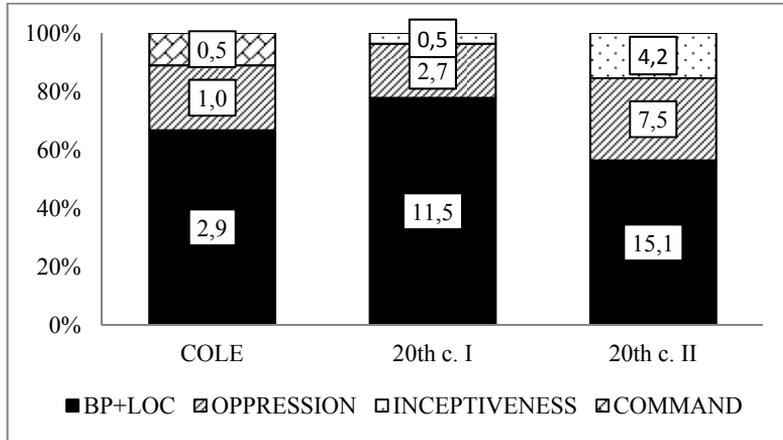
Figure 66. The distribution of the functions of *jalge all* (feet+under), instances pmw

The data show that not all of the functions that are used in contemporary data are present in all periods, and that there are functions no longer used in contemporary language. Namely, in COLE there is one example (0.5 instances pmw) where *jalge all* (feet+under) is used to express the meaning 'in one's command' (see example (284)). Such a usage is not present in contemporary Estonian. However, the same meaning can be expressed using the phrase *jalge ette* (feet+before). Owing to their frequency of use and context, such expressions are idiomatic rather than representing complex function words. As this is a sole example, it is not discussed any further.

(284) *... Jssand,   sa   olle-d   innimes-t   pan-nud   wallitse-ma*
    Lord   you   be-2SG   human-PRT   put-PST.PTCP   rule-SUP
    omma-ø   kät-te   töe-ø   ülle;   keik   olle-d   sa
    own-GEN   hand-PL.GEN   work-GEN   over   all   be-2SG   you
    *temma-ø   **jalg-e**   **alla**   pan-nud: puddolojusse-d*
    s/he-GEN   foot-PL.GEN   under.LAT   put-PST.PTCP   animal-PL
    ja   härja-d,   keik   puhhas,   ja   ka   metsalojusse-d: linnu-d,
    and   bull-PL   all   entirely   and   also   wild animal-PL   bird-PL
    mis   taewa-ø   al,   ja   kalla-d,   mis   mere-s,   mis
    what   sky-GEN   under.LOC   and   fish-PL   what   sea-INE   what
    mere-ø   kõhha-d   läbbikäi-wad.
    sea-GEN   place-PL   go round-3PL
    'Lord, you have placed man to rule over your handiwork; you have put
    everything under his feet: domestic animals and oxen, all, and also wild animals;
    birds that are under the sky and fish that are in the sea, that go round the places
    in the sea.' [COLE, Masing, 1818, 108]

264

Figure 66 also indicates that the function INCEPTIVENESS is not present in COLE. It does not appear until the 1st half of the 20th century, where it is represented by 0.5 instance pmw (3%). In the 2nd half of the 20th century, the amount of examples that express this function increase to 4.2 instances pmw (16%). The data suggest that the relative number of examples expressing OPPRESSION is more stable over the observed periods. Although the relative frequency increases (from 1.0 to 2.7 to 7.5 instances pmw), the proportion among the other functions remains between 18%–28% in all the observed periods.

Thus, the data suggest that the function INCEPTIVENESS may be a more recent development than OPPRESSION, as it appears later in the data and increased in frequency during the following periods. However, as the data sampel is small, it is also possible that the function was already present in COLE, but did not occur in the corpus due to its low frequency of use. The examples belonging to the simple structure are more frequent than the complex structure in all of the observed periods.

4.8.7.2. *Jalge all* (feet+under) as a complex adverb and a complex postposition

In contemporary Estonian, *jalge all* (feet+under) is mostly used as a complex adverb. However, complex postpositional uses also occur, but infrequently (see section 4.4). The available diachronic data suggests that in the case of *jalge all* (feet+under), adverbial use is more common than postpositional use in most decades. Moreover, adverbial uses increase over time, reaching 96% in the 2nd half of the 20th. It was also observed in the synchronic analysis that *jalge all* (feet+under) is mostly used as a complex adverb in contemporary language (see section 4.4). The diachronic data suggests that even though the first instances of *jalge all* (feet+under) in the 19th century as a complex unit include both – adverbial as well as postpositional uses, it is only the adverbial uses that keep increasing over time, making up 66% by the 1st half of the 20th century and over 90% of the usages as complex units in the 2nd half of the 20th century. Thus, the available data does not suggest a directed change towards more grammatical usages among the complex units. Rather, it seems that the complex postpositional uses have been there as a marginal trait throughout the observed periods.
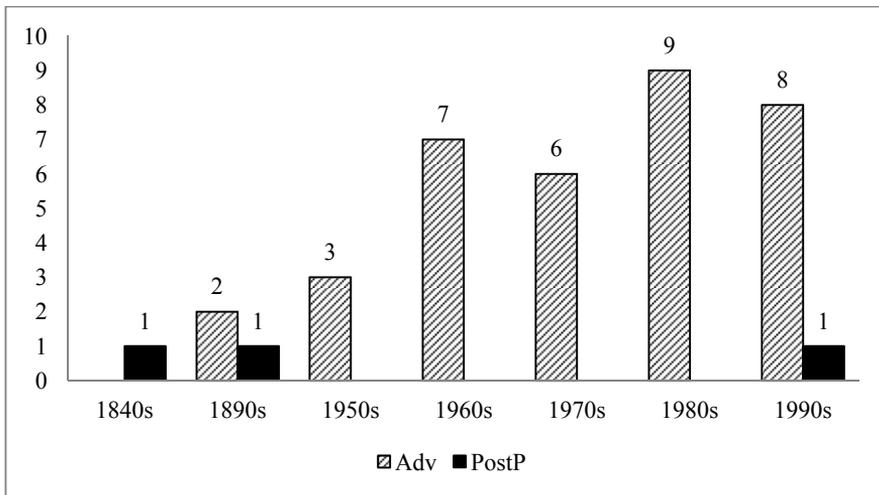
Figure 67. The first attestations of *jalge all* (feet+under) as a complex unit in the diachronic data in absolute frequencies per decade

The data does not allow me to conclude which part-of speech appeared first. The data show that the adverbial and postpositional uses appear in the same century. The first complex postpositional use appears in 1840s (see example (285)) and the first example of adverbial use appears in 1890s (see example (286)). Although the adverbial use appears half a century later in the data, there are too few examples available to conclude that adverbial uses did not exist at that time.

(285) *Kõik   mailm   nägg-i   ja   teäd-i-s   kui   otsata   sure*
       all   world   see-PST.3SG and   know-PST-3SG   how   awfully   large
       wäe-ga   temma   Wenne maa-Ø   sísse   lä-k-s,   et   se-dda
       force-COM   s/he   Russia-GEN   into   go-PST-3SG   to   this-PRT
       omma-Ø   jalg-e   álla   tálla-ta,   ja   kuida
       own-GEN   foot-PL.GEN under.LAT   trample-INF   and   how
       lubba-s,   omma   wäesuur-te-le   jagga-da,   ja   nüid
       promise-PST.3SG   own-PST   general-PL-ALL   share-INF   and   now
       pidd-i   temma   sure-Ø   häbbi-ga   kui   üks   näljane
       must-PST.SG3   s/he   great-GEN shame-COM   as   one   hungry
       rebbane   koer-te-ga   wälja   ásse-tud   sa-ma.
       fox   dog-PL-COM   out   drive-PST.PTCP   get-SUP
       Lit. All the world saw and knew how awfully large was the army that he took to
       Russia in order to trample it under his feet and how he promised to share the loot
       with his generals; and now he was like a hungry fox to be driven away by the dogs.
       'All the world saw and knew how awfully large was the army that he took to
       Russia in order to tread on it and how he promised to share the loot with his
       generals; and now he was like a hungry fox to be driven away by the dogs.'
       [COLE, Suve Jaan, 1841, 83]

266

(286) *Sakslase-d   wõi-wad   parema-d   inimese-d   oll-a,   kui   meie,*
German-PL   might-3PL   better-PL   human-PL   be-INF   than   we
aga   nende-l   ei   ole-Ø   õigus-t,   mei-d
but   they-ADE   NEG   be-NEG   right-PRT   we-PL.PRT
oma-Ø   looma-de-ks   alandada,   mei-d   palja-ks
own-GEN   animal-PL-TRL   degrade-INF   we-PRT   poor-TRL
riisu-da   ja   kõik-i   hä-i-d   omadus-i   meie-Ø
rob-INF   and   all-PL.PRT   good-PL.PRT   quality-PL.PRT   we-GEN
*sees   **jalge   alla**   tallata ...*
in   feet   under   trample
Lit. The german may be better than us but they have no right to treat us as their
animals, rob us blind and trample all our good qualities under their feet.
'The Germans may be better than us but they have no right to treat us like animals
animals, rob us blind, and tread uponour good qualities' [AJA1890\pro0005]

Thus, both parts-of-speech were already present in the 19th century but the order
of the development of the adverbial and postpositional functions remains incon-
clusive. There are also no instances of hybrid forms in the diachronic data, i.e.
jalge all (feet+under) does not occur with adjectival PNs at all. Very few (only
1 out of 1918) instances of such usages can be found in etTenTen, so adjectival
PNs are extremely rare in contemporary usage (see section 4.2).

As *jalge all* (feet+under) was seldom used as a complex postposition, it is
quite expected that there is not much variation among the PNs in diachronic
data. In this context the diachronic data only includes PNs that refer to animate
beings – humans and animals (see examples (287)–(288) respectively).

(287) *... sügisjaheduse-st   kõvakstõmbunud   rohi   krudise-s*
autumn coldness-ELA   stiff turned   grass   crunch-PST.3SG
*ta-ø   **jalg-e   all,**   ning   metsavaht,   vaatamata*
s/he-GEN   foot-PL.GEN   under.LOC   and   forester   in spite of
oma-ø   vigasuse-le,   otse   kihele-s   meelehea-st
own-GEN   cripple-ALL   straight   tingle-PST.3SG   liking-ELA
ja   elurõõmu-st.
and   sparkle-ELA
'… the grass that had turned stiff from the autumn coldness crunched under his
feet and the forester, in spite of his crippleness, downright tingled of happiness
and joy of life.' [ILU1936\ram0030]

(288) *Aga   pea   wisa-t-i   te-da   hobus-te*
but   soon   throw-IMPS-PST   s/he-PRT   horse-PL.GEN
***jalg-e   alla,**   ja   sealt   ühe-ø   põlewa-ø*
foot-PL.GEN   under.LAT   and   there   one-GEN   burning-GEN
puu-ø   riida-ø   peale
tree-GEN   stack-GEN   on
'But soon s/he was thrown under the horses' feet and from there onto a burning
stack of wood.' [AJA1890\pro0029]

However, as the synchronic analysis revealed (see section 4.5.1.6) that collective PNs are present in contemporary usage as a marginal trait, it could be suggested that the use of *jalge all* (feet+under) has extended. However, as discussed in see section 4.5.1.6, such usages are not considered as grammatical, but rather as fixed expressions. Thus, the data suggest that the use of PNs of *jalge all* (feet+under) when used as a complex postposition has not extended. Decategorialization could not be analyzed in the data of *jalge all* (feet+under), because in this case categorialization was not observable in terms of non-agreement.

4.8.7.3. Summary the diachronic analysis of *jalge all* (feet+under)

The results of the diachronic analysis confirm the conclusion drawn based on contemporary data, i.e. *jalge all* (feet+under) is the least grammaticalized of the studied phrases. As the use of *jalge all* (feet+under) entails few of the parameters of grammaticalization, the diachronic data did not reveal much evidence of increasing grammaticalization.

The first attestation of *jalge all* (feet+under) appears in the corpus in 1818. In this example, *jalge all* (feet+under) is used as a freely combined unit. However, the immediate context (verb lemma) suggests this usage is a suitable bridging context for the development of the function OPPRESSION, which appears as a complex function. The data show that the latter are also already present in the 19th century, but their relative frequency does not increase over the investigated periods. However, the function INCEPTIVENESS does not occur in the data of COLE, but is present in the data for the 1st part of the 20th century and its frequency in the data increases from then on. Thus, the amount of complex units is also increasing in the data during the observed periods. However, usage of freely combined phrases of *jalge all* (feet+under) remains more prominent in all of the observed periods (including the contemporary data).

The data suggest that similar to the contemporary data, historically adverbial uses are more common than postpositional uses among the complex units. As there is very little data on complex postpositional uses, it is difficult to draw conclusions regarding contextual expansions regarding the PN. However, the data does suggest that *jalge all* (feet+under) was not used with adjectival PNs during the earlier periods. As adjectival PNs are extremely rare in contemporary usage, this is quite expected. The data also suggests that in the 19th and 20th centuries, *jalge all* (feet+under) only occurred with animate PNs. In the contemporary data, there are also a few examples where *jalge all* (feet+under) co-occurs with a PN that refers to a collective. However, such examples do not represent usages as complex postpositions, but rather as idiomatic expressions. Thus, this extension cannot be taken to suggest grammaticalization.

4.8.8. Main conclusions based on the diachronic analysis

The aim of the diachronic analysis was to gain an insight into the earlier stages of grammaticalization and be able to describe the grammaticalization paths of the body part related complex function words. Thus, in the analysis of the diachronic data, the following questions were addressed:

1. Is there evidence of increasing grammaticalization?
2. What is the diachronic order of the changes?
3. What motivates the changes?

First of all, it must once more be stated that all of the conclusions made based on the diachronic data must be taken with caution because in the case of most of the phrases there was not enough data available to be able to properly observe the development of the studied phrases, and because the texts from different periods do not belong to the same genres. However, as the diachronic corpora are the only available electronic collections of texts that allow me to observe the studied phrases in different periods, it is essential that this data was analyzed.

In order to observe if the studied phrases are becoming more grammatical, the following aspects and their dynamics were observed – their relative frequency in the observed periods, the distribution of the freely combined phrases and complex units among the studied phrases, (where appropriate) the distribution of the usage functions, the semantic classes of the PNs, and the agreement between the first component of the phrase and its PN. In the following, I will summarize the most important results of these aspects, and draw conclusions based on the analysis of the diachronic data.

The data suggest that the relative frequency of all the phrases, except *kaela peal* (neck+on), increases over the observed period. *Kaela peal's* (neck+on) decrease may have been due to it having been replaced by a developing simple function word *kaelas* (neck+INE). The fact that most of the phrases are used increasingly over the studied periods is consistent with principles of grammaticalization. However, as an increase in frequency is predominantly associated with grammaticalizing items, it makes sense to observe the frequency of the phrases among the free and complex structures separately.

Regarding the distribution of the freely combined phrases and complex units, the data suggest an increase in most cases – *käe all* (hand+under), *selja taga* (back+behind), *külje all* (side+under), *käe kõrval* (hand+beside) and *jalge all* (feet+under). However, *käe kõrval* (hand+beside) and *kaela peal* (neck+on) have had a different development. In the case of *käe kõrval* (hand+beside), it appears that the complex units have always been more common, and only in the later periods (20th c. II and etTenTen) include a few examples of the freely combined phrases. In the case of *kaela peal* (neck+on), the frequency of complex units has decreased since the period of Old Literary Estonian. Although complex units may be over represented in COLE, it is still clear that *kaela peal* (neck+on) used to be a fairly established complex unit which has

since been replaced by other expressions. Thus in this aspect, it seems that the development of complex units is better observable among the phrases *käe all* (hand+under), *selja taga* (back+behind), *külje all* (side+under) and *jalge all* (feet+under).

The analysis of the semantic classes of the PNs mostly shows increasing diversity. The data suggest that the semantic classes have expanded among the phrases *käe all* (hand+under), *selja taga* (back+behind) (except for the function 'hidden by') and *külje all* (side+under). However, in the case of the phrases *kaela peal* (neck+on), *jalge all* (feet+under) and *käe kõrval* (hand+beside), such dynamics is not clearly observable.

Regarding the agreement functions of the phrases, it can be observed that non-agreement is perhaps the most varied parameter among the different phrases. The data suggest that non-agreement develops in the later periods (2nd half of the 20th century) in the case of *käe all* (hand+under). Plural forms appear later in the case of *käe kõrval* (hand+beside). However, in the latter case, the appearance of plural forms might not suggest extension, but rather be a consequence of different genres or the vast differences in the sample sizes (i.e. were possible earlier, but did not appear in the corpus). In case of *selja taga* (back+behind), the frequency of non-agreement fluctuates over the observed periods. There are also differences between the functions. For some, non-agreement increases over the centuries, for others (e.g. COVERTNESS) it decreases. In the case of some phrases, the non-agreement shows no change in dynamics over the observed periods. In the case of *kaela peal* (neck+on), the proportion of plural forms remained more-or-less the same between COLE and etTenTen. In the case of *külje all* (side+under), the amount of plural forms remained low. As the synchronic analysis showed that non-agreement is present in both functions in contemporary language, the diachronic analysis confirms these results.

In most cases, hybrid forms are not present in the diachronic data, but do occur as a marginal trait in the contemporary data. Only in the case of *käe all* (hand+under), which was used as part of a larger idiomatic expression in the earlier periods, does the data suggest that the frequency of hybrid forms decreases throughout the centuries. Thus, it seems that the possibility to use an adjectival PN is more dependent on the specific phrase, rather than connected to a stage of grammaticalization.

Regarding the role of the adverbial in the developmental path of the phrases, the data suggest that each phrase and in some cases each usage function, has its own path. Some phrases (the spatio-temporal use of *selja taga* (back+behind)) go through the adverbial intermediate stage before they become complex postpositions, as suggested by the model of Habicht and Penjam (2007). Some phrases (*käe all* (hand+under) and *selja taga* (back+behind) (when used to express CONCEALMENT), and *kaela peal* (neck+on)) seem to skip the adverbial stage and develop straight to complex postpositions. However, in some cases the chronological order of the adverbial and postpositional stages remains

inconclusive, because there is not enough data amongst the earlier uses to decide whether the phrase was not used as both parts of speech (e.g. *jalge all* (feet+under)). In some cases, the adverbial and postpositional usages appear temporally so close that it is impossible to determine whether one of them developed earlier (i.e. for *külje all* (side+under) and *selja taga* (back+behind) (in the function SUPPORT). The latter may also point to the fact that the adverbial and postpositional usages appeared more or less at the same time and neither was prior to the other. It is possible that once the components of the phrase are analyzed as a complex unit, i.e. the phrase is lexicalized, it can be instantly used as both a complex adverb and a complex postposition.

Possible foreign influence on the development of Estonian complex function words was observed in two cases – *käe all* (hand+beside) and *kaela peal* (neck+on). Both of these items may have developed due to a loan translation from German. In the case on *käe all* (hand+under), the foreign influence is more probable because the instances where foreign influence is observed stand right next to its possible source in the corpus. It is quite probable that the Estonian expression (*end) Jumala vägeva käe alla alandama* 'to humble (oneself) under the mighty hand of God' are word-for-word translations of the German expression *unter seine gewaltige Hand.* However, as these are also the first (and only) instances of *käe all* (hand+under) during the earlier period, and there is some evidence of extension of this particular expression to other contexts, it is possible that Estonian *käe all* (hand+under) has its roots in this religious formula. There was no such evidence for *kaela peal* (neck+on). Nevertheless, the Estonian-German dictionary of Wiedemann (1973 [1869]: 178), translates the phrase as *Einem auf dem Hals liegen, zur Last sein* (lit. to lie on one's neck; 'to burden somebody'). Thus, because the two languages were in close contact during this time, it is possible that that this meaning might have developed due to (or by support of) German influence.

Nevertheless, the diachronic data does not answer the most important question – how did the other complex function words came to be. Because there is very little data from the earlier periods, the very beginning of the process remains unrevealed. However, the analysis shows that almost all the phrases were already interpretable as holistic units from the earliest examples. Moreover, in some cases, the data indicates that the holistically interpretable cases appeared earlier or were more frequent than the freely combined phrases. A similar observation has been made by Hoffmann when analyzing the diachronic development of *by way of*. Hoffmann claims that the data suggests as if *by way of* behaved like a complex preposition from its early attestations in that it expressed abstract functions, did not display formal variation, rarely occurred with a determiner, and displayed other (syntactic) features that would suggests its advanced level of grammaticalization. On the other hand, his data suggests that more recently, concrete usages of *by way of* have began to emerege suggesting a reversed type of change that starts from abstract and ends up in concerte usages. Hoffmann concludes that while *by way of* is an untypical

instance of grammaticalization, the possibility of the reversal development needs further investigation (Hoffmann 2005: 69–71). In the case of Estonian data, however, it is rather the lack of data which does not allow me to observe the development of complex postpostions from an earlier period. Thus, even though the grammaticalization is considered to be in incipient, the very beginning of it has not been caught on record. However Mair (2011: 248–249) suggests that the 'incipient grammaticalization' is an ambigupus status. When investigating the conjunction usage of *on basis of*, the claims that linguistic items may function in the same grammatical item-like stage for centuries and not develop towards a more grammatical usage or it is also possible that a change that has started a long time ago is 'waiting' for an impulse to launch. Thus, it is possible that the phrases lexicalized long time ago are now gradually making their way towards more grammatical usages.

V CONCLUSION

5.1. The aim of the study

The present thesis focuses on the development of complex function words in Estonian – complex adverbs and complex postpositions. Although the development of complex function words has not been the central focus of grammaticalization studies, there is a considerable body of literature available on the topic. However, so far, the discussion has mainly focused on the development of complex *pre*positions, and fewer studies can be found on complex *post*positions. The present study aims to contribute to the ongoing debate on complex adpositions with data from Estonian, which is predominantly a postpositional language. Most of the previous literature on complex prepositions has described the phenomenon in Indo-European languages, e.g. in English (Hoffmann 2004, 2005; McMichael 2006; Smith 2013); German (Lehmann 1998; Trawiński 2003; Rostila 2004, 2006; etc.), Spanish (Lehmann 2002), French (Adler 2008), Dutch (Moirón, Bouma 2003), and Swedish (Sigurd 1993). The present study offers an overview of the phenomenon in Estonian, a language that belongs to a different family (Uralic languages) and is also typologically different from those mentioned above.

Because of the typological differences between the languages and structural differences between pre- and postpositions, the process of development of the complex adpositions as well as the criteria for determining the status of the complex adposition are different. Thus, most of the syntactic tests that have been applied to determine the status of the complex preposition in Indo-European languages (for an overview of approaches to the English complex prepositions see Hoffmann 2005: 26 ff), cannot be used when studying the development of complex postpositions in Estonian. Therefore, one of the goals of this study was to develop a methodology to describe and study complex postpositions.

Studying the development of Estonian complex postpositions enables me to take closer look at two topics of interest – the early stages of grammaticalization and the relationship between grammaticalization and lexicalization. First, the development of complex function words results in a new grammatical item – a postposition. Therefore, this phenomenon is taken to be an instance of grammaticalization. Because the rise of complex postpositions is a relatively recent development, it is considered to be an instance of incipient grammaticalization. Thus, observing this phenomenon involves looking closer at the very beginning of the grammaticalization process. This may help gain some insight into the earliest stages of the grammaticalization of function words (complex as well as simple), which may not be observable in retrospect. Second, the development of complex postpositions is considered to be an instance of grammaticalization which also includes lexicalization. In this study, grammaticalization and lexicalization are not considered to be opposite phenomena but complementary processes that can be combined in a single instance of language

change. The presence of both processes in the development of complex post-positions allows me to further investigate the role and effect of lexicalization in the process of grammaticalization.

The source form of the complex function word is the simple postpositional phrase. In contemporary Estonian, source and target forms exist hand in hand. For instance, in (289) *selja taga* (back+behind) expresses its compositional meaning 'at a location posterior to one's back', and is, therefore, considered to illustrate usage as a simple postpositional phrase. In (290), on the other hand, the phrase expresses an abstract holistic, i.e. non-compositional, meaning and, as such, is analyzed as a complex adverb. However, in actual language use, we find instances where the phrases carrying abstract, non-compositional meaning are used in sentential contexts where they can be (re)analyzed as postpositions. This is exemplified in (291), where the relation between the landmark of the phrase containing *selja taga* (*muutused* 'changes') and the trajector (in this case: schematic) is expressed by *selja taga* (back+behind) as a whole. Thus, in this context, the complex unit *selja taga* (back+behind) is analyzed as a postposition.

(289) *Iga-ks   juhu-ks   roni-s   ta   oma-Ø   voodi-st   siiski*
      each-TR   case-TR   climb-PST.3SG   s/he   own-GEN   bed-ELA   yet
      välja   ja   puge-s   vaikselt   Lilli   voodi-sse,   õe-Ø
      out   and   ceep-3PST.3SG   quietly   Lilli   bed-ILL   sister-GEN
      *kaitsva-Ø   **selja-Ø**   **taha.***
      protective-GEN   back-GEN   behind.LAT
      'Just in case, s/he climbed out of his/her bed and crept silently to Lilli's bed behind her sister's protective back' [www.poogen.ee]

(290) *Kujutlusvõime   maali-b   pilt-e   eeloleva-st*
      imaginaton   paint-3SG   picture-PL.PRT   preceding-ELA
      *ähvardava-st   hukkumise-st   ning   **seljataga**   varitseva-st*
      treathening-ELA   perishing-ELA   and   back.behind.LOC   lurking-ELA
      orjapõlve-st   või   surma-st.
      slavehood-ELA   or   death-ELA
      'The imagination paints pictures of the threatening perishing ahead and the slavehood or death lurking behind the back.' [www.advent.ee]

(291) *Ja   ärge   puge-ge   maailmamajanduse-Ø   **muutus-te***
      and   NEG   hide-IMP.2PL   world economy-GEN   change-PL.GEN
      selja-Ø**   **taha   *kui   te   mitte midagi   ei   suuda-Ø*
      back-GEN   behind.LAT   if   you   nothing   NEG   can-CONNEG
      ette näh-a   ja   isegi   selge-i-d   protsess-e   riigi-s
      forsee-INF   and   even   clear-PL-PRT   process-PL.PRT   state-INE
      tunneta-da,   siis   pan-ge   oma-Ø   pilli-d   kotti-Ø.
      feel-INF   then   put-PL.IMP.2PL   own-GEN   instrument-PL   bag-ILL.
      Lit. And do not hide behind the back of the changes in world economy if you cannot foresee anything and even feel the clear processes in the state then quit.
      'And do not hide behind the changes in world economy if you cannot foresee anything and even feel the clear processes in the state then quit.' [www.epl.ee]

The present study proposes a way to classify a group of items (such as *selja taga* in example (290)) which are currently incompatible with the traditional view of Estonian language structure. Although Estonian adverbs may be either simple or complex (see example (291) above), the traditional view does not acknowledge the existence of complex postpositions in Estonian. Most Estonian grammars do not mention complex postpositions at all; according to Palmeos (1985: 6), the category of postpositions includes simple members only. Nevertheless, she admits that it is possible for some complex adverbs to 'carry the prepositional function' (ibid.).This thesis demonstrates that Estonian has complex function words which behave as postpositions in certain contexts. Moreover, I claim that in some cases, the complex function words arise first as complex postpositions and develop the adverbial function only later.

The scope of this study is restricted thematically. The study is concerned with body part related postpositional phrases – *selja taga* (back+behind), *käe all* (hand+under), *külje all* (side+under), *käe kõrval* (hand+beside), *kaela peal* (neck+on), and *jalge all* (feet+under) – whose selection criteria and process is described in section 3.1. This is by no means considered to be an exhaustive list of grammaticalizing postpositional phrases. As can be observed in the previous literature, the development of complex function words in Estonian is not confined to body part related phrases.

As there is not much previous literature on this topic, the main question of the present study is whether Estonian has complex postpositions. The idea that Estonian is developing a (sub)category of complex postpositions was first suggested by Habicht and Penjam (2007) and has been further developed in Jürine (2011), and Jürine, Habicht (2013). However, these studies are based on linguistic experiments and/or observations, and the question has not been investigated with a corpus. The present account represents a corpus linguistic approach, making use of a large corpus (etTenTen), compiled of 270,000,000 words, which is representative of contemporary Estonian. A large-scale corpus analysis provides enough data for an exhaustive study of the phenomenon at hand. In addition, making use of corpora allows me to observe text frequency as well as pattern frequencies of the object of study. The development of complex postpositions has so far only been studied from the synchronic perspective. However, as grammaticalization is a process which evolves gradually over time, the phenomenon should also be studied from a diachronic perspective. Thus, in addition to the synchronic analysis, the present account also takes a diachronic approach, through which I observe the development of complex function words over time.

The corpus study seeks answers to the following research questions:

1. What evidence supports the claim that the freely combined simple postpositional phrases have been reanalyzed as complex structures?

2. What is the role of frequency in the development of complex function words? What evidence is there of fixedness and productivity of the phrases under investigation?
3. Is there diachronic evidence of (further) grammaticalization in the development of complex postpositions?
4. Is adverbialization a vital prerequisite of the development of complex postpositions in Estonian?
5. What is the role of lexicalization in the process of grammaticalization of complex postpositions in Estonian?
6. Is the development of Estonian complex function words a language-internal development or a contact-induced change?

5.2. The results of the synchronic and diachronic analysis

1. What evidence supports the claim that the freely combined simple postpositional phrases have been reanalyzed as complex structures?

Based on previous research, the general principles of grammaticalization and lexicalization, and observational inspection of the object of study, I propose a set of criteria to distinguish between the simple and complex structure and observed the spread of these features in the synchronic and diachronic data. The criteria characterizing the complex structure are: **unit interpretation, extension beyond human reference**, and **non-agreement of the body part term and the preceding (pro)noun** (i.e. decategorialization). The present study also considers **frequency**, which has often been associated with grammaticalization. In this account, absolute as well as pattern frequency is used to observe **fixedness** and **productivity** of the phrases in question. In the following, these criteria will be discussed one by one.

One of the most prominent features of the complex structure is unit interpretation. In order to analyze a phrase as a complex adverb or complex postposition in any given example, it is mandatory that in this example, the phrase is interpreted as a unit, i.e. it carries a holistic meaning. A holistic meaning is here understood as a meaning that is not the sum of the meanings of its components and, as such, it agrees with the definition of lexicalization by Brinton and Traugott (2005: 96). The same principle underlies the developmental process of Estonian complex adverbs, which have most often developed from simple postpositional phrases, e.g. *selja taga* (back+behind) when expressing a temporal meaning. Thus, unit interpretation is considered to be one of the key criteria in the development and identification of a complex unit. Once a postpositional phrase has developed a holistic meaning, it becomes a potential candidate for becoming a complex function word.

However, the status of complex postposition also requires that a compulsory structural criterion – the immediate adjacency of the postposition and its (pro)nominal complement – is met. Thus, only instances of lexicalized phrases that meet this criterion can be analyzed as complex postpositions. For instance, usages like the one illustrated in (292) are analyzed as complex postpositions but the usage illustrated in (293) is not because the presence of the adjective *hoolitsev* 'caring' violates the criterion of immediate adjacency. Usages exemplified in (293) are not very common in my data but, nevertheless, present in all the studied phrases. Because such examples have characteristics of both structures, they can be viewed as in-between cases of the simple and the complex structure, and they are referred to as 'hybrid forms'. The existence of similar usages in other languages have been suggested to indicate that reanalysis is not an abrupt but a gradual change (cf. De Smedt 2012). Although the hybrid forms of the phrases under investigation here are likely examples of utterances that are being reanalyzed, the present object of study does not provide a suitable material to investigate the nature of reanalysis because there are no conflicting *formal* features that would suggest belonging to two structures at the same time.

(292) *Mina*   *ole-n*   *AINULT*   *tarka-de*   *ja*   *toreda-te*
     I   be-1SG   only   smart-PL.PRT   and   nice-PL.PRT
     arsti-de   ***käe-Ø***   ***alla***   *sattu-nud* …
     doctor-PL.PRT   hand-SG   under.LAT   happen-PST.PTCP
     Lit. I have only chanced under the hand of clever and nice doctors …
     'I have only chanced under care of clever and nice doctors …'
     [rahvahaal.delfi.ee]

(293) *Ära*   *muretse-Ø..*   *haigla-s*   *ole-d*   ***arsti-de***
     NEG   worry-CONNEG   hospital-INE   be-2SG   doctor-PL.GEN
     hoolitseva-Ø   ***käe-Ø***   ***all***   *ja*   *räägi-Ø*   *kindlasti*
     caring-GEN   hand-GEN   under.LAT   and   tell-IMP.2SG   surely
     ära,   *et*   *p-ole*   *ümber*   *keera-n*
     away   that   NEG-be   around   turn-PST.PTCP
     'Don't worry… In hospital you will be under the caring hand of the doctors; and make sure to tell them that [the baby] has not turned around, etc.'
     [www.nupsu.ee]

The corpus analysis suggests that the distribution of holistic and compositional usages varies among the phrases investigated. The proportion of holistic usages is the largest (86%–99%) in the case of *käe all* (hand+under), *käe kõrval* (hand+beside), and *külje all* (side+under). In the case of *selja taga* (back+behind), *jalge all* (feet+under) and *kaela peal* (neck+on), the proportion of lexicalized and compositional usages remains around 50%.

As suggested above, unit interpretation is indicative of lexicalization. However, as postpositions are considered to be grammatical items, the development of complex postpositions, on the other hand, is considered to be an instance of

grammaticalization. Therefore, the question arises of what criteria can be found that are indicative of grammaticalization. In the following, I discuss the criteria that suggest that the lexicalized usages of these phrases have been reanalyzed as complex postpositions, which are grammatical items.

Desemanticization. The data suggest that the lexicalized meanings are more abstract than the compositional meanings, which refer to locations relative to body parts. This is in accordance with their interpretation as function words which express relations (locations, time, manner, etc.) rather than making reference to concrete entities, such as human body parts. As such, the lexicalized usages of the body part related phrases are indicative of desemanticization because in such cases the referential capacity of the body part terms is fading. Desemanticization is considered to be one of the key parameters of grammaticalization (Heine, Kuteva 2002, 2007). As the source forms lose their semantic content, they express various abstract functions, such as SPACE/TIME, PROXIMITY, CONTROL, and others. In most cases, one phrase expresses several functions. Also in most cases, these functions can be analyzed as different stages on the same developmental path. For instance *käe all* (hand+under) expresses PHYSICAL CONTROL as well as MENTAL CONTROL. However, in some cases there seems to be no apparent link between the different functions of the same phrase. For instance, *selja taga* (back+behind) expresses SPACE/TIME, CONVERTNESS, SUPPORT, and CONCEALMENT which are considered to be different instances of lexicalization of the source form.

The functions carried by the phrases are connected to the meanings of the source form. In some cases, the semantic shifts are rather expected, because the results converge with findings from other languages. For instance, expressions involving the body part terms *külg* 'side' or *käsi* 'hand' have come to express PROXIMITY or CONTROL in many languages. Nevertheless, in some cases the functions covered by the phrases are more specific, i.e. the phrases have gone under desemanticization, but still semantically more contentful than the cases described above. For instance, as a complex unit, *kaela peal* (neck+on) expresses BURDEN, which is clearly an abstract concept, especially relative to the compositional meaning of the phrase. However, BURDEN is also a specific enough function not to constitute a widely spread semantic shift. The specificity is also taken to suggest that, in such functions, the phrases are grammaticalized to a lesser degree because grammaticalization is taken to be a process towards greater abstractness and schematicity. This applies especially for *jalge all* (feet+under), which expresses OPPRESSION and INCEPTIVENESS. Such specific usages do not suggest typological generality, and are as such more characteristic of lexical items and not (yet) grammatical items.

Extension. Because of the unit interpretation and desemanticization, the usages that pertain to the complex structure may occur in new sentential contexts. Heine and Kuteva (2002, 2007) consider extension to new contexts to be one of the crucial parameters of grammaticalization. In this study, extension is

observed in the semantic class of the (pro)nominal that precedes the body part related phrase (PN). In the case of the freely combined phrase, the PN behaves as the modifier of the body part term, while in the case of the complex postposition the PN behaves as the complement of the complex postposition. When analyzed as complex units, the phrases may take PNs that refer to (human) collectives, institutions and, in a few individual cases, even to abstract notions. This is considered to suggest extension because such usages are semantically incompatible with the compositional meaning – the collective noun cannot normally modify a body part term. For instance the utterance *toimetuse käe all* (lit. under newsroom's hand) makes sense as a complex unit, but not as a freely combined phrase.

It is quite expected for the inanimate PNs to refer to (human) collectives and institutions. It is clear that the usages with collective PNs have evolved in the context of human PNs. It is also likely that the PNs referring to institutions are the result of the development towards more abstract uses, whereby the body part related phrase is further dissociated from its literal meaning. The abstract PNs are very rare, and at this point it is difficult to explain their connection with the rest of the semantic classes. However, it is clear that they represent the most abstract uses of the studied phrases and, therefore, in such cases, the link with the body part term is considered to be the weakest.

Usages indicative of extension are not particularly common among any of the studied phrases. The data show that extension is most widely spread in the case of *selja taga* (back+behind) and *kaela peal* (neck+on), which are the phrases of highest and lowest frequency in my data. Thus, there seems to be no correlation between the overall frequency of the phrase and the frequency of usages that involve extension. However, as there are very few examples (N = 70) where *kaela peal* (neck+on) is used as a postposition, it is possible that the collective PNs are overrepresented. The data indicates that the collective, institutional, and abstract PNs that were considered to be incompatible with the simple structure are indeed only used with complex postpositions. Nevertheless, a few exceptions were also observed. In extremely rare cases (1%; 20 examples out of 2178), collective PNs were also found among the phrases that were analyzed as freely combined phrases. However, such examples may also be considered in-between cases that may be analyzed as either simple or complex structure. The presence of the inanimate PN suggests the latter interpretation.

Thus, the data show that the PNs referring to collectives, institutions and abstract notions predominantly occur with the complex structure. Even though extension is not particularly widespread, based on its distribution among the simple and complex structure and the analysis of the semantic classes, it is concluded that the presence of such usages suggests extension. The extension of the complex postpositional uses is considered to indicate actualization of the reanalysis.

Decategorialization. In contemporary Estonian, the PNs of the phrases under consideration may occur in both singular and plural form. The plural form

is regarded to be semantically incompatible with the simple structure. Thus, lack of agreement in number between the PN and the body part term is considered to be semantically incompatible with the case of the freely combined phrases but suitable for the complex structure. Hence, non-agreement suggests that the phrasal structure has been reanalyzed.

The data show that both singular and plural PNs are present in all of the phrases[71] but similarly to extension, the non-agreement is not widely attested in my data. The data showed that the proportion of examples showing non-agreement varies from 3% to 31%. Non-agreement is most common among examples of *kaela peal* (neck+on), where it appears in 31% of the examples. The plural PNs also appear in a considerable proportion of examples of the phrases *käe all* (hand+under) (21%), *selja taga* (back+behind) (18%), and to a lesser extent in the case of *käe kõrval* (hand+beside) (12%). Plural PNs are seldom used with the phrase *külje all* (side+under) where they make up only 3% of the examples.

The data show that non-agreement may occur with both the simple and the complex structure. The analysis indicates that most of the phrases (except for *külje all* (side+under)) are more likely to exhibit non-agreement with the complex structure. The preference for the complex structure is particularly clear in the case of *kaela peal* (neck+on), where non-agreement is present in 41% of the examples with the complex postposition, and in 5% of the freely combined phrases. These results are also statistically significant, and the effect size is intermediate. With the other phrases, the correlation between the non-agreement and the complex structure is significant but weak. Thus, it is also possible that the significance is due to a relatively large sample. It is possible that the non-agreement does not suggest decategorialization but is rather an example of free variation. However, it is also possible that it does indicate decategorialization, and, therefore, reanalysis, but as the grammaticalization of complex postpositions is still in its early stages, the distribution of plural forms is not yet clear. Thus, non-agreement is not as clear a criterion as extension beyond human reference, and therefore, its possible role in this type of grammaticalization requires more investigation.

---

[71] Except for *jalge all* (feet+under) where the non-agreement is not observable. In this case, the body part term *jalad* refers to a body part that comes in pairs, and is plural in this phrase to begin with. The singular PN would not suggest semantic incompatibility either, because the phrases *lapse jalad* (child.GEN feet 'child's feet') and *las-te jalad* (child-PL.GEN feet 'children's feet') are equally possible, while *õpetaja-te käsi* (teacher-PL.GEN hand 'teachers' hand') is not.

2. What is the role of frequency in the development of complex function words? What evidence is there of fixedness and productivity of the phrases under investigation?

Making use of corpora and data on real language use allows me to observe the role of frequency in the development of complex function words. In addition to observing the text and pattern frequency of the studied phrases, more sophisticated methods for measuring the strength of association between the words was implemented – mutual information (MI) and log-likelihood measure. MI was used to measure the collocational strength between the components of the studied phrases. Log-likelihood was used to measure the collocational strength between the studied phrases and other elements in the immediate sentential context. The former is taken to suggest fixedness and the latter is taken to show productivity of the studied phrases.

Fixedness. The analysis shows that the studied phrases are rather fixed. MI scores of the studied phrases were analyzed in relation to that of the phrases with a plural first component (*selja taga* (back+behind) vs. *selgade taga* (backs behind)) and to that of the other freely combined body part related postpositional phrases that are logically possible in language external reality (*nina kõrval* 'beside one's nose') but which do not have a unit interpretation. As expected, the scores of the studied phrases were considerably higher than those of compositional phrases as well as the phrases with plural body part terms. However, the absolute frequencies of body part related phrases with a plural first component were quite low in general. These results suggest that the studied phrases are more tightly bound units than the freely combined postpositional phrases and that they are freezing in the singular form. This is taken to suggest that the studied phrases are becoming autonomous complex units.

On the other hand, it is possible to observe the association scores of the studied phrases relative to each other. As the studied phrases vary in their absolute frequencies (from 216 to 10,958 instances of occurrence), it is to be expected that the scores of the association measures do not suggest an equally strong association between all of the studied phrases. In general, the results of the MI analysis comply with the ranking based on absolute frequency of the phrases. That is, the most frequent phrase *selja taga* (back+behind) yielded the strongest association between its components, followed by phrases of intermediate frequency, *külje all* (side+under*), jalge all* (feet+under) and *käe all* (hand+under); phrases with the lowest absolute frequency – *käe kõrval* (hand+beside) and *kaela peal* (neck+on) – also exhibit the lowest MI score. Thus, it is taken that the more frequent phrases are likely to be more tightly bound as complex units.

Productivity. Collocational strength was also measured between the complex units and their immediate sentential context – PN and verb lemmas. The strongest collocates of each phrase as a complex unit were determined by the log-likelihood measure. Examples with the strongest collocates of each phrase

were not considered to be formed productively. Thus, productivity was measured as the proportion of weaker collocates of the complex units.

While the results show that each phrase has its own set of strong collocates, they also suggest that the more frequent phrases are used rather productively. The usages with the stronger collocates are mostly just more typical examples of the complex postpositions, not idiomatic expressions. Either way, the strongest collocates are considered to suggest formulaic use, and, therefore, are not considered to be formed productively. The results show that based on their use with (pro)nominal and verbal complements, *käe all* (hand+under) and *külje all* (side+under) can be considered most productive and *kaela peal* (neck+on), and especially *jalge all* (feet+under) are less productive. Although *jalge all* (feet+under) is intermediately frequent phrase in the present data set, the analysis suggest that the contexts of *jalge all* (feet+under) are especially restricted – the productive uses of the verb remain below 10% in case of both of its functions as a complex unit. As high productivity is usually associated with grammaticalization, the more frequent, abstract and tightly bound phrases can be considered to be more grammatical. The phrases that are contextually restricted are considered to be less grammatical.

3. Is there diachronic evidence of (further) grammaticalization in the development of complex postpositions?

Before summarizing the results of the diachronic analysis, it must be underlined that because the data are few, the conclusions must be taken with caution. However, diachronic analysis was considered necessary because the available data enables insight into the diachronic development of the phrases under investigation and reveals aspects that are not observable in the contemporary language. Thus, in the diachronic analysis the development of complex function words was observed in three periods – the 17th to 19th centuries, 1st half of the 20th century and the 2nd half of the 20th century.

The analysis shows that the overall frequency of the phrases increased in most cases over the observed periods. The only exception is *kaela peal* (neck+on), which was highly frequent in Old Literary Estonian but the frequency fell abruptly in the early 20th century. I suggest that *kaela peal* (neck+on) may be losing ground to its morphologically simple rival *kaelas* (neck+INE) which is also a rising function word which bears a similar function. The diachronic data of *kaela peal* (neck+on) helps to explain the results of its contemporary analysis, which suggest that while the phrase is the least frequent of the included phrases, it is the most advanced with regard to extension and decategorialization. Thus, it may be the case that *kaela peal* (neck+on) was more grammaticalized in Old Literary Estonian but was replaced by its morphologically simpler rival. The diachronic data also show that, in most cases, the relative frequencies of the complex units were increasing over the observed

periods. An exception is the phrase *käe kõrval* (hand+beside), which has a high, stable proportion of complex units in all the periods.

The diachronic data do not include many examples of extension. In most cases they only appear in the later periods. It is possible that the usages with collective/institutional and abstract PNs are not present in the diachronic corpora because such usages are also quite rare in contemporary data which comes from a much larger corpus. However, it is equally possible that such usages were simply not possible during these periods, and can only be found in the contemporary data. The non-agreement, which is taken to suggest decategorialization, is more widespread in the diachronic data. However, in this criterion, the variation between the phrases was the broadest. In some cases, non-agreement appears in the later periods (*käe all* (hand+under)), in some cases (*selja taga* (back+behind)) it is observable in all the periods. In some cases, the relative frequency of examples with non-agreement increases, while in others it does not. As in contemporary language, non-agreement appears with both structures, but it is difficult to make any general conclusions based on the diachronic data. We can conclude from the diachronic analysis that in most cases, overall frequency as well as relative frequency of the studied phrases increases. The diachronic data also yield valuable insight into the development of the individual phrases, which is also touched on in answering the remainder of the research questions.

4. Is adverbialization a vital prerequisite of the development of complex postpositions?

The analysis indicates that the use of the adverbial and postpositional function varies between the phrases and is dependent on the semantics of the particular complex item. The diachronic analysis suggests that both directions of change are possible: in the case of some phrases (e.g. *käe kõrval* (hand+beside)), the data suggests a proportional increase in adverbial uses; in other cases (e.g. *selja taga* when expressing SUPPORT) we see a proportional increase in postpositional uses. In some cases the proportions of complex adverbial and postpositional uses remain the same. As the postpositions are considered to be more grammatical than adverbs, the increase of the postpositional uses would suggest an increase in grammatical uses. However, based on my data, this does not seem to be the case among all the phrases.

The same applies to the diachronic order of the adverbial and postpositional uses. The data suggest that, although the adverbial intermediate stage is needed in some cases, it is not a crucial step in the developmental path of the complex postpositions. For instance, in the case of *selja taga* (back+behind) expressing the function CONCEALMENT, the data show that the complex unit was only used as a postposition at first, and that adverbial uses have emerged as a marginal feature only in the contemporary data. The available data indicates a similar path for *käe all* (hand+under) and *kaela peal* (neck+on).

5. What is the role of lexicalization in the process of grammaticalization of complex postpositions in Estonian?

In the present account, lexicalization of the phrases is not considered to be directly associated with adverbialization, but rather with the adoption of the complex item into the mental lexicon (cf. Brinton, Traugott 2005). As mentioned above, the data show that the adverbial stage does not necessarily precede the postpositional usage. However, I maintain that lexicalization is a vital intermediate stage in this type of grammaticalization process.

Lexicalization is considered to be vital in this type of grammaticalization process because the development of complex postpositions constitutes an atypical kind of grammaticalization, where the grammaticalizing item is a multi-word unit, not a single word or a word form. Because the phrase enters the grammaticalization process as a holistic unit, it is vital that the phrase develop a unit interpretation, i.e. is first lexicalized. Secondly, the analysis shows that unit interpretation is the single most reliable criterion for distinguishing the freely combined phrases and the complex units. Although the complex unit *may* exhibit extension beyond human reference and decategorialization, in order to be analyzed as a complex unit, the phrase simply *must* be lexicalized. For instance, (294) is a typical example of *kaela peal* (neck+on) used as a complex postposition, but it does not exhibit either extension or decategorialization; but it is lexicalized. It is possible that once the components of the phrase are analyzed as a complex unit, i.e. the phrase is lexicalized, it can then be used as either – the complex postposition or the complex adverb.

(294) *Ja   niisuguse-d   inemese-d   nagu   Ramo,   Liiv   ja   sarnase-d*
      and   this kind-PL   people-pL   like   Ramo   Liiv   and   like-PL
      *"targa-d"   laiskvorsti-d   istu-vad   **minu-Ø   kaela-Ø   peal**.*
      smart   lazybones-PL   sit-3PL   I-GEN   neck-GEN   on.LAT
      Lit. And people like Ramo, Liiv and similar "smart" lazybones sit on my neck.
      'And people like Ramo, Liiv and similar "smart" lazybones depend.on me'
      [arhiiv.koolielu.ee]

The preliminary stage of lexicalization affects the grammaticalization process in that it enables larger semantic shifts at a time. This is not meant to say that lexicalization represent an abrupt change. To the contrary, the development of complex units is taken to be a gradual process (cf. Brinton, Traugott 2005) because the development of phrasal meanings occurs slowly over time. However, lexicalization enables the phrases to develop rather abstract meanings without any intermediate stages or bridging contexts that would be expected in typical instances of grammaticalization. For instance, *selja taga* (back+behind) expresses many rather abstract functions (COVERTNESS, SUPPORT, CONCEALMENT), all of which are considered to be separate instances of lexicalization. While their connection with the source meaning is evident, they do not form a single grammaticalization path with gradual transitions, but rather

they form quite discreet individual usage patterns. In addition, there is unusually little formal evidence of grammaticalization, such as instances of extension and decategorialization. Thus, it is taken that lexicalization has provided the necessary conditions for the phrases to be reanalyzed as complex units, but the process of grammaticalization, which takes place through the gradual spread of these complex items in the language system, is still in its early stages.

6. Is the development of Estonian complex function words a language-internal development or a contact-induced change?

The diachronic data do not suggest foreign influence in the case of most of the phrases included in the study. However, with two of the phrases – *käe all* (hand+under) and *kaela peal* (neck+on) – German influence is suspected. In the case of *käe all* (hand+under), the German influence is indicated by the available parallel text in German, where the expression *unter seine gewaltige Hand* 'under his mighty hand' is used. For *kaela peal* (neck+on), a similar expression in German (*einem auf dem Hals liegen* 'to lie on one's neck') is found in Wiedemann's Estonian-German Dictionary (1973) [1893], pointing to the possibility of contact-induced change. Nevertheless, as these German expressions are not grammatical items, but rather idiomatic expressions, the foreign influence is considered to result in loan translations rather than grammatical copying. Of course, the mere fact that no evidence of foreign influence was found in the case of the other phrases does not exclude the possibility of contact-induced change. Nevertheless, the fact that most of the phrases are listed in Wiedemann's Dictionary (1973) [1893] and do not have word-for-word translations suggests that these phrases were lexicalized by the latter 19th century, and that German influence is unlikely.

5.3. Evaluation of the methodology and directions for future research

In the present study, the development of complex postpositions in Estonian was investigated synchronically and diachronically. The synchronic analysis is based on etTenTen, a large (270,000,000-word) corpus compiled of texts collected from Estonian web pages. The diachronic analysis is based on data from the Corpus of Old Literary Estonian (COLE), 19th-century texts, and the Corpus of Estonian Literary Language (CELL). In the following I present a critical evaluation of the methods and data sources used.

etTenTen may be considered an appropriate source of data for the synchronic analysis of complex function words. The corpus is large enough to provide an adequate amount of data for the analysis of more frequent phrases as well as less frequent ones. The data represent different types of texts, covering

genres closer to spoken language (e.g. blog entries) as well as those closer to standard language (e.g. official web sites). As such, it provides a data sample that reflects the language change studied here representatively. However, the comparison of the etTenTen data with that of earlier periods potentially yields some methodological problems. For instance, CELL, which covers the 20th century, consists of fiction and journalistic texts. As such, it only represents standard language (at least starting from the 1920s, the beginning of the language planning era). Because etTenTen also includes texts with unedited and nonstandard language use, it is not certain whether differences in the results between the corpora are indicative of language change or change of genre. The same caveat applies when comparing the results from the 17th to 19th centuries (COLE + 19th century texts) and CELL. The former periods, especially up to the 19th century, include mostly texts pertaining to the religious sphere. In addition, the periodization (17th to 19th century, 1st first half of the 20th century, and 2nd half of the 20th century) may not reflect the language change accurately because these periods cover unequal amount of time (half a century vs. centuries). However, as limited data is available, this proved to be the best way to observe the development of the phrases diachronically. Despite the caveats mentioned above, the diachronic analysis enables valuable insight into the earlier periods of the development of complex postpositions. It also provides answers to questions regarding the diachronic development of the complex grams (e.g. the diachronic order of the adverbial and postpositional function, as well as German influence on the development of *käe all* (hand+under) and *kaela peal* (neck+on)).

One of the goals of this study was to develop a methodology to distinguish between the simple and complex structure. As suggested above, the most prominent criterion for distinguishing between the freely combined phrases and complex units is the semantic interpretation of the phrase. In addition to the semantic criteria, two formal features are indicative of actualization of the reanalysis of the phrases – semantic classes of the (pro)noun and non-agreement of the preceding (pro)noun and the first component of the phrase.

The nouns that refer to collectives, institutions, and abstract notions are predominantly found among the instances where the phrases are analyzed holistically. Even though extension beyond human reference is not a vital feature for the phrase to be analyzed as a complex unit, the distribution of the semantic classes suggests that such usages co-occur with the complex structure, almost without exception. Thus, the semantic class of the preceding (pro)nominal is considered a suitable criteria to distinguish between the simple and the complex structure.

Furthermore, observing the distribution of the semantic classes enables me to trace the developmental paths of the complex items. In most cases the only semantic class present on both structures is human beings. It is most likely that the complex postpositions have developed in the context of animate (human) PNs. Thus, the (human) collective and institutional (pro)nouns are a step closer

to more abstract uses, which has been made possible by the semantic change that must have taken place in the context of animate PNs. Such semantic shift is expected in the case of some functional categories (e.g. MENTAL CONTROL, PHYSICAL CONTROL (*käe all* (hand+under)), COVERTNESS (*selja taga* (back+behind)), (mental) BURDEN (*kaela peal* (neck+on)), because these notions are associated with human behavior. However, with phrases that also express spatial notions, such as PROXIMITY (*külje all* (side+under)), or SPACE(/TIME) (*selja taga* (back+behind)), a different kind of developmental path would be expected. It has been suggested (Svorou 1994, Heine 1997) that spatial function words that have developed from body part terms go through the stage of referring to object part. However, this is not a likely scenario in the case of *selja taga* (back+behind), because the data suggest that instances where *selja taga* (back+behind) was used to express a location relative to an object part were quite rare. Moreover, the spatio-temporal function is mostly used with human complements, so there seems to be no connection between the complex locative function word and object part.

The exclusion of the object part stage from the developmental path is connected to the fact that the body part terms in most of the phrases are not pro-ductively used to refer to object parts in Estonian. However, there are two phrases in my data – *külje all* (side+under) and *kaela peal* (neck+on) – in which the first component is used productively to refer to object parts. As was suggested above, *kaela peal* (neck+on) expresses mental burden and is, accordingly, used in the context of humans (with the simple as well as the complex structure). However, the phrase *külje all* (side+under) is used in reference to humans as well as objects, and the data did not reveal which con-text has been more conducive to the development of the sense of proximity. It may be a parallel development in both contexts. This is also suggested by the fact that the complex unit occurs with complements that refer to collectives and complements that refer to regions. As suggested above, the former are likely to have developed from human complements and the latter from (natural) objects.

In future research, the extension of complex units could also be observed in other elements besides the preceding (pro)noun. For example, it would be in-teresting to observe the semantic class of the LM in the case of complex adverbs to see if semantic classes other than collective, institution, and abstract notion are possible, and how common these classes are.

As stated above, the data show that non-agreement may occur with both the simple and the complex structure. Although it seems that non-agreement prefers the complex structure, it is still less discreetly divided between the analytically and holistically analyzed phrases than the semantic classes of the preceding pronoun. Plural and singular forms are especially evenly distributed among *selja taga* (back+behind) and *külje all* (side+under). These phrases express functions closer to the simple structure than the other studied phrases – *külje all* (side+under) expresses PROXIMITY and *selja taga* (back+behind) expresses spatio-temporal function (among other functions). Hence, it may be expected

that other features of their usage (e.g. non-agreement) are less discreetly distributed as well. In case of the phrases *käe all* (hand+under), *käe kõrval* (hand+beside) and *kaela peal* (neck+on), the plural forms were preferred with complex units. However, as the effect size did not (in most cases) suggest a strong correlation between the plural forms and the complex units, more research is needed in order to specify the role of non-agreement in the development of complex postpositions. It is possible that non-agreement does not suggest (only) decategorization but is associated with an archaic trait of Estonian, which allows to refer to certain vovabulary (including plural body parts) formally in singular. This trait has receded but is preserved in some idiomatic expressions. (Alvre 1968, Õim, Õim 2015)

The results related to frequency – absolute frequency, fixedness, and productivity – were rather consistent among the phrases. That is, the highly frequent phrases were also more fixed and highly productive. The only exception was the phrase *jalge all* (feet+under), which was intermediately frequent among the phrases in the study, but rather unproductive as a complex unit, appearing as a complex postposition quite infrequently. However, the frequency of the phrases was not necessarily connected with the other observed features – the distribution of the semantic class and grammatical number of the preceding pronoun. The occurrence of these features seems to be rather dependent on the concrete phrase and its literal meaning and function(s). It is also possible that these features are still too infrequent in general to observe any preference toward more or less frequent phrases. It is also possible that the number of phrases investigated is too small. Thus, the correlation of frequency and other features suggestive of grammaticalization needs some further investigation in a data sample that includes more phrases.

5.4. Concluding remarks

This study shows that body part related postpositional phrases exhibit certain features that are indicative of their use as complex items – complex postpositions and complex adverbs. According to the traditional view, the category of Estonian postpositions involves only simple items. Thus, it is demonstrated that the Estonian language is developing a new subcategory of function words – the subcategory of complex postpositions. In the contemporary language, the source and target forms exist side by side. It is suggested that the source and target forms, i.e. the simple and the complex structure are distinguished primarily based on their meaning. The complex items are lexicalized, or analyzed holistically, and as they are used to express abstract functions, they are also desemanticized. Some formal evidence also supports the claim that the phrases have been reanalyzed as complex postpositions – the phrases exhibited extension and to a certain extent decategorialization.

In this study, it is demonstrated that certain body part related postpositional phrases are not compatible with the traditional view on Estonian language structure, and that these expressions are analyzable as complex postpositions. Drawing on the general principle of grammaticalization and lexicalization, an initial methodology was developed to determine which elements belong to the subcategory of complex postpositions. This methodology can be tested and further developed in future studies on grammaticalization of complex post-positions.

VI SUMMARY IN ESTONIAN

6.1. Uurimisobjekt ja teema põhjendus

Siinne töö keskendub liitsete funktsioonisõnade – adpositsioonide ja adverbide – kujunemisele eesti keeles. Kuigi traditsiooniliste käsitluste järgi sisaldab eesti keele adverbide klass nii liht- kui ka liitvorme, on adpositsiooni peetud vaid lihtüksusi sisaldavaks sõnaklassiks. Harva esineva joonena on liitseid adpositsioone eesti keeles siiski võimalikuks peetud, näiteks nendib Palmeos, et mõned liitadverbid (nt -*poole*, -*pool*, -*poolt* järelkomponendiga üksused) esinevad liitprepositsioonide funktsioonis (Palmeos 1985: 6). Eesti keeles langevad adpositsioonid ja adverbid leksikaalsel tasandil suures osas kokku, st samad sõnad võivad käituda nii adverbi kui ka adpositsioonina (*pane kampsun alla* (*alla* kui adverb), *pane kampsun jope alla* (*alla* kui postpositsioon). See viitab adverbide ja adpositsioonide ühisele arenguteele – tüüpiliselt tekib funktsioonisõna adverbina ja hakkab seejärel täitma ka postpositsiooni funktsioone. Siinses töös lähtutakse eeldusest, et kui lihttüvelised funktsioonisõnad esinevad nii adverbi kui ka adpositsioonina, on sarnane kasutuskontekstide laienemine võimalik ka liitüksuste puhul. Töö eesmärk on arutleda liitsete adpositsioonide, täpsemalt liitpostpositsioonide võimalikkuse üle eesti keeles ning pakkuda välja semantilised ja morfosüntaktilised kriteeriumid, mille alusel selliseid liitüksusi määratleda. Üldisema eesmärgina võib nimetada ka liitsete grammatikaüksuste tekkimise ja määratlemise problemaatika tutvustamist sellise keele põhjal, kus domineerivad postpositsioonid. Töö loodab panustada ka üldisemasse liitadpositsioonide teemalisse arutellu, mis on seni puudutanud peamiselt indoeuroopa keeli ja keskendunud põhiliselt liitprepositsioonidele (vt nt Sigurd 1993; Lehmann 1998; Moirón, Bouma 2003; Rostila 2004; Hoffmann 2005; Smith 2013).

Siinse käsitluse kohaselt on eesti keeles liitsete funktsioonisõnade allikvormiks lihtkaassõnafraas. Allik- ja sihtvormid eksisteerivad tänapäeva keeles kõrvuti. Näiteks väljendab *selja taga* näites (295) otsest, st kompositsioonilist tähendust, milleks on inimese suhtes tagapool paiknemine, ning on seega analüüsitud lihtkaassõnaühendina ehk lihtstruktuurina. Näites (296) väljendab *selja taga* aga abstraktse(ma)t terviktähendust, st on mittekompositsiooniline, ja on sellisena analüüsitud liitadverbina. Tegelikus keelekasutuses leidub ka selliseid juhte, kus abstraktset tervikutähendust väljendavad üksused esinevad liitpostpositsiooni funktsioonis. Sellist kasutust illustreerib näide (297), kus liitsele terviküksusele *selja taga* eelneb sellega ühte fraasistruktuuri kuuluv orientiir (*muutused*), mille suhet trajektooriga (siin: väljajäetud pronoomen) väljendab *selja taga* tervikuna. Seega käitub *selja taga* selles näites liitpostpositsioonina. Nii liitadverbi kui ka liitpostpositsiooni kohta kasutatakse töös terminit *liitstruktuur* (ingl *complex structure*).

(295)   *Igaks juhuks ronis ta oma voodist siiski välja ja puges vaikselt Lilli voodisse,*
          *õe kaitsva **selja taha**.* [www.poogen.ee]

(296)   *Kujutlusvõime maalib pilte eelolevast ähvardavast hukkumisest ning **seljataga***
          varitsevast orjapõlvest või surmast. [www.advent.ee]

(297)   *Ja ärge pugege maailmamajanduse **muutuste selja taha** - kui te mitte midagi ei*
          suuda ette näha ja isegi selgeid protsesse riigis tunnetada, siis pange oma
          pillid kotti. [www.epl.ee]

Teema on eesti keeleteaduse seisukohalt oluline, kuna aitab klassifitseerida selliseid üksusi, millel ei ole praeguses keelestruktuuri kirjelduses kindlat kohta. Töös pakutakse siin analüüsitava materjali põhjal välja kriteeriumid, mille alusel eesti keele liitpostpositsioone määratleda. Kuna eesti keeles seostub iga tüüpi liitsõna staatus automaatselt ortograafilise liitsõnaga, saab töö tulemusi arvesse võtta ka eesti keele kokku- ja lahkukirjutamise reeglite sõnastamisel. Praegu kehtiva normingu järgi saab eesti standardkeeles *selja taga* kokku kirjutada näites (295), kuid mitte näites (296), sest reegel tõlgendab samasse fraasistruktuuri kuuluvat noomenit (siin: *muutused*) alati kehaosanime täiendina. Ent kuna *selja taga* on siin leksikaliseerunud, st väljendab iseseisvat abstraktset terviktähendust, on kõnealust noomenit kohasem tõlgendada liitpostpositsiooni laiendina. Kui *muutused* ei ole sõna *selg* täiend ja *selja taga* moodustab fraasis tihedamini kokkukuuluva üksuse, ei saa lahkukirjutamise nõuet sellega põhjendada. Habicht ja Penjam (2007) on välja toonud, et kaassõnade norminguvastane kokkukirjutamine on eesti keele toimetamata (ja mõnel määral ka toimetatud) tekstides silmatorkav problem. Jürine (2011) on täheldanud seost kokkukirjutamise ja kaassõnafraasi tähenduse tajumise vahel. Kuna siinses uurimuses kokkukirjutamist eraldi kriteeriumina ei vaadelda, nimetatakse liitpostpositsioonideks (*complex postpositions*) kõiki üksusi, mis täidavad oma funktsioonilt ja struktuurilt eesti keele postpositsiooni kriteeriume (vt allpool).[72] Siinses töös ei välistata, et adpositsioonifraaside norminguvastane kokkukirjuta-mine on üks võimalik tunnus, mis viitab liitsete funktsioonisõnade tekkimisele. Siiski jäi see aspekt siinsest käsitlusest välja, kuna tegemist on spetsiifilise ortograafilise tunnusega, mis on üksuste lingvistilise analüüsi seisukohast sekundaarne.

    Liitsete funktsioonisõnade uurimine ei ole mõistagi ainult eesti keele seisukohalt oluline küsimus. Kuigi liitadpositsioonide tekkimine ei ole olnud grammatisatsiooniuuringute keskne teema, leidub uurimusi, mis käsitlevad liitkaassõnade tekkimist ja arenguradu kui grammatisatsiooniprotsessi (vt nt

---

[72] Jürine ja Habicht (2012) on teinud vahet liitsel postpositsioonil ja liitpostpositsioonil, esimene neist viitab ülal kirjeldatud liitse struktuuriga üksustele, mis käituvad nagu post-positsioonid, kuid mille kirjapilt seda ei kajasta; liitpostpositsioonidena käsitlevad nad eespool kirjeldatud üksusi, mis moodustavad ka ortograafilise sõna. Siinses käsitluses kasu-tatakse terminit liitpostpositsioon (*complex postposition*) nende mõlema kohta, st kokku- ja lahkukirjutistel ei tehta sisulist vahet.

Lehmann 1998, 2002; Rostila 2004; Hoffmann 2005). Kuna mitmesõnalised üksused, mis käituvad keelekasutuses terviküksustena, on probleemiks keele automaatsel töötlemisel, leidub liitadpositsioonide määratlemise teemal ka formaalsemaid ja praktilisema eesmärgiga uuringuid (nt Sigurd 1993). Enamik nimetatud uuringutest käsitleb liitprepositsioone, mitte liitpostpositsioone, nagu käesolev uurimus. Niisiis, siinne uurimus püüab täita lünka liitsete adpositsioonifraaside tekkimises ja kirjeldamises. Kuigi liitprepositsioonide ja liitpostpositsioonide tekkimine on olemuselt sarnane protsess, on neis fraasistruktuurist tulenevaid erinevusi, mis kajastuvad ka liitüksuste määratlemise kriteeriumides. Kui kõrvutada eesti keele liitpostpositsioone inglise keele liitprepositsioonidega, võib sarnaste joontena välja tuua selle, et mõlemas keeles tekivad liitsed adpositsioonid lihtadpositsioonifraasist. Näiteks fraasid struktuuriga PNP (prepositsioon-nimisõna-prepositsioon), nagu *in view of*, mis tänapäeva inglise keeles kannab abstraktset tähendust ('silmas pidades, valguses') on samas ka lahutatav lihtüksusteks (vrdl eesti keele näidetega (295) ja (297) eespool).

Samas pakub inglise keele struktuur selliste üksuste tuvastamiseks vormilisi kriteeriume, näiteks artikli kadu (*in view of* vs. *in the view of*), kuid eesti keeles sellised formaalsed kriteeriumid ei rakendu. Siinse uurimuse üks eesmärke ongi töötada välja kriteeriumid, mis eristaksid eesti keele lihtkaassõnafraasi ja liitpostpositsiooni. Neid kriteeriume tutvustatakse kokkuvõtte alapeatükis 6.4.1.

6.2. Materjal ja meetod

Uuritava materjali moodustavad kehaosanimetustega seotud kaassõnaühendid, st fraasid, mis koosnevad kehaosanimest (nt *selg*, *käsi*) ja lihtkaassõnast (nt *all*, *kõrval*). Fraasid valiti välja kehaosanimetuste ja ruumikaassõnade võimalike kombinatsioonide hulgast, mille komponendid on loetletud Lisas 1 (Annex 1).

Uuritavate fraaside valikul kehtisid järgnevad kriteeriumid:
– fraas kannab terviktähendust, st on leksikaliseerunud,
– fraasi kasutus ei piirdu tänapäeva keeles väga kitsaste kontekstidega,
– fraas esineb kontekstides, mis võimaldavad selle reanalüüsimist liitpostpositsiooniks,
– semantiline nihe on piisavalt suur, et eristada liht- ja liitstruktuuri,
– leksikaliseerunud kasutused on piisavalt sagedased, et need võimaldaks analüüsida fraasi kasutust liitüksusena.

Kriteeriumide alusel valiti kehaosanimetuse ja lihtkaassõna võimalike ja loogiliste kombinatsioonide hulgast tänapäeva keele seisukohast kõige sobivamad fraasid, milleks olid **selja taga**, **käe all**, **külje all**, **käe kõrval**, **kaela peal** ja **jalge all**. Mõistagi ei ole see nimekiri ammendav, rohkemal või vähemal määral vastavad nendele kriteeriumidele ka muud kehaosanimetustega seotud kaassõnafraasid ja ka muu tähendusega fraasid. Siin uuritav nähtus ei piirdu

ainult kehaosanimetusi sisaldavate fraasidega, vaid esindab laialdasemat tendentsi (vt Habicht, Penjam 2007; Jürine 2009; Jürine 2011).

Siinne uurimus võtab aluseks kasutuspõhise lähenemise põhimõtted: analüüs põhineb tegeliku keelekasutuse näidetel. Tegemist on korpusuuringuga, mille materjal pärineb põhiosas eesti internetikorpusest etTenTen[73]. Korpus sisaldab 270 000 000 sõnet ning tekste 686 000 eestikeelselt veebilehelt, mis esindavad järgmistesse valdkondadesse kuuluvaid tekste: valitsus, foorumid, religioon, blogid, perioodika, teabekirjandus, klassifitseerimata. Sellisena esindab korpus suulise ja kirjaliku keele vahepealset žanri/registrit, sest sisaldab nii standard-keelele kui ka suulisele (argi-) suhtlusele lähedast keelekasutust. Kuna keele-muutused kajastuvad tihti enne suulises keeles või tekstiliikides, mis ei esinda standardkeelt, sisaldab korpus sobivat materjali siinse uurimuse eesmärgi täitmiseks.

Uurimus on põhiosas sünkrooniline, sest eesti keele liitpostpositsioonide teket võiks pidada üsna hiljutiseks arenguks, kuna selleteemalised uurimused peaaegu puuduvad (v.a Habicht ja Penjam 2007; Jürine 2011; Jürine, Habicht 2013). Senistes uurimustes on vaadeldud eelkõige liitpostpositsioonide selliseid aspekte, mida on võimalik uurida vaid sünkrooniliselt (nt kaassõnaühendite kokkukirjutamine tänapäeva (toimetamata) kirjalikus keeles). Grammatisatsioon on oma olemuselt aeglaselt kulgev protsess, mida tavaliselt jälgitakse diakrooni-lises analüüsis. Seega arvestatakse ka siinses uurimuses diakroonilist materjali, vaadeldes uuritavaid fraase ja nende arengut alates nende esimestest esine-mustest 17. sajandi kirjakeeles kuni 1990ndate aastateni ning võrreldes neid tulemusi etTenTenil põhineva sünkroonilise analüüsi tulemustega.

Siinsele tööle annab metoodilise aluse grammatisatsiooniteeoria ning uurita-vate fraaside analüüs lähtub grammatisatsiooni üldtunnustatud põhimõtetest. Kesksel kohal on Heine ja Kuteva (2002, 2007) sõnastatud grammatisatsiooni parameetrid: desemantiseerumine, ekstensioon, dekategoriseerumine ja erosioon. Siinse nähtuse puhul on siiski tegemist veel üsna algses faasis oleva gram-matisatsiooniprotsessiga ning seega ei saa eeldada, et nimetatud parameetrid on uurimismaterjalis väga laialdaselt esindatud, eriti mis puudutab grammatisat-siooniprotsessis toimuvaid hilisemaid muutusi – näiteks erosiooni kui foneetilise ainese kadumist.[74] Lisaks põimitakse käsitlusse teisi grammatisat-siooniga (ja üldisemalt keele muutumisega) seostatavaid mehhanisme, nagu reanalüüs ja selle aktualiseerumine. Kuna liitkaassõnade tekkeprotsess kätkeb endas ka komplekssete üksuste teket, hõlmab see endas ühtlasi leksikali-seerumist, st grammatisatsiooni ja leksikalisatsiooni käsitatakse siin koos kulgevate nähtustena.

---

[73] http://www2.keeleveeb.ee/dict/corpus/ettenten/about.html (Kasutatud 11.01.2016)
[74] Selgeid märke lühenemisest või häälikute kaost uuritavate kaassõnaühendite puhul ei ole ning kuna siinne töö keskendub nähtuse uurimisele kirjakeeles, ei ole foneetilisi kriteeriume siinsesse käsitlusse kaasatud.

Siinne käsitlus liitpostpositsioonide arengust tugineb suuresti Habichti ja Penjami (2007) kirjeldatud eesti keele funktsioonisõnade tsüklilise arengu mudelile. Mudeli kohaselt tekivad eesti keele funktsioonisõnad arengutsüklis, kus leksikaalsest üksusest ja grammatilisest komponendist (nimisõna ja käände-lõpp, nt *jooks*+ADE) tekkinud lihttüvelised funktsioonisõnad – liitadverbid ja liitkaassõnad (*jooksul*) – on tänapäeval haaratud uude grammatisatsiooni-tsüklisse, kus uued funktsioonisõnad tekivad taas leksikaalse ja grammatilise üksuse põimumisel. Liitsete funktsioonisõnade tekkimisel on leksikaalseks üksuseks e allikvormiks aga nimisõna ja lihtkaassõna ühend.

6.3. Uurimisküsimused

Siinses doktoritöös otsitakse vastuseid järgmistele küsimustele:
1. Milliseid tõendeid leidub selle kohta, et kaassõnaühendeid kasutatakse terviküksustena, mis esinevad adverbi ja/või postpositsiooni funktsioonis, st milliste kriteeriumide alusel eristada liht- ja liitstruktuuri?
2. Missugune on sageduse roll liitsete funktsioonisõnade arengus? Kuivõrd kokkukuuluvaks ja produktiivseks võib uuritavaid fraase pidada?
3. Milliseid tõendeid leidub uuritavate liitsete funktsioonisõnade diakroonilise arengu kohta? Mis ajast pärinevad uuritavate fraaside esimesed kasutus-juhud liitüksustena? Kas aja jooksul on märgata liitüksuste kasutussageduse osakaalu suurenemist kõigi kasutusjuhtude seas? Kas uuritavatel perioodidel ilmneb märke grammatisatsioonile viitavate faktorite levimisest?
4. Kas liitüksused, mis esinevad tänapäeva keeles nii adverbi kui ka post-positsioonina, on kujunenud adverbistumise ja seejärel kaassõnastumise teel (Habicht, Penjam 2007) või on mõeldav ka vastupidine arengutee?
5. Missugust rolli mängib leksikaliseerumine liitpostpositsioonide tekke-protsessis?
6. Kas uuritavad liitpostpositsioonid on tekkinud loomuliku keelemuutuse käigus või on tegemist kontaktimõjulise keelemuutusega?

Järgnevalt esitatakse sünkroonilise ja diakroonilise uurimuse tulemused vastustena püstitatud uurimisküsimustele.

6.4. Tulemused ja arutelu

6.4.1. Sünkroonilise analüüsi tulemused

1. Milliseid tõendeid leidub selle kohta, et kaassõnaühendeid kasutatakse terviküksustena, mis esinevad adverbi ja/või postpositsiooni funktsioonis, st milliste kriteeriumide alusel eristada liht- ja liitstruktuuri?

Materjali esmasel vaatlusel töötati välja grammatisatsiooni üldpõhimõtetele tuginevad kriteeriumid, mille alusel eristada lihtstruktuuri (allikvorm) ja

liitstruktuuri (sihtvorm). Seejärel vaadeldi nende kriteeriumide levikut uuritavas materjalis. Liitüksustele võib iseloomulikuks pidada järgmisi tunnuseid: liitüksused kannavad terviktähendust, mis on tekkinud ühendi leksikaliseerudes; liitüksused on levinud kasutuskontekstidesse, mis pole lihtüksuse puhul võimalikud või esinevad väga harva (ekstensioon); liitüksused kalduvad minetama oma morfosüntaktilisi omadusi, mis on omased vabalt kombineeruvale fraasile (dekategoriseerumine). Järgnevalt selgitatakse neid kriteeriume ükshaaval lähemalt.

Ühe olulisema kriteeriumina võib nimetada terviklikku tähendust. Selleks et kaassõnafraasi saaks käsitada liitadverbi või liitpostpositsioonina, on eelduseks, et sellel on tekkinud terviktähendus. Terviktähenduse all mõeldakse siin seda, et fraas kannab tähendust, mis ei ole otseselt tuletatav selle komponentide tähendustest (nt *selja taga* tähenduses 'ajaliselt) möödas' – *Raske päev oli seljataga*). Sellisena vastab fraaside terviktähenduse kujunemine leksikalisatsiooni definitsioonile (Brinton, Traugott 2005). Ka eesti keele liitadverbide tekkimist, mille allikvormiks on sagedasti kaassõnafraasid (Kasik 2013), on kirjeldatud protsessina, kus kaks sõna "on sulandunud kokku ühtseks määrsõnaliseks väljendiks" (Erelt et al. 1995: 597). Terviktähenduse olemasolu (*unit interpretation*) on liht- ja liitüksuste eristamisel kriitilise tähtsusega. Kui kaassõnafraas väljendab terviktähendust, mis ei ole selle komponentide summa, on tegu potentsiaalse liitse funktsioonisõnaga. Uuritavaid fraase tõlgendati siinses töös liitüksuseks, kui need väljendasid terviktähendust ja kui nende struktuur täitis liitkaassõnale seatud kriteeriumi, milleks on kaassõna ja seda laiendava noomeni vahetu kõrvuti paiknemine, sh adjektiivse täiendsõna vahele asetamise võimatus. Seetõttu analüüsitakse kasutusjuhte, mida illustreerib näide (298), liitpostpositsioonideks, sest sellisel juhul on täidetud mõlemad tingimused, kuid kasutusjuhud, mida illustreerib näide (299), liitpostpositsiooniks ei kvalifitseeru, sest potentsiaalse postpositsiooni ja selle laiendi (siin: *arstide*) vahel paikneb adjektiiv *hoolitsev*.

(298) *Mina olen AINULT tarkade ja toredate **arstide käe alla** sattunud ja soovin kõigile tublidele Eesti arstidele edaspidiseks jõudu ja tervist kõigepealt iseendale ja siis kannatlikkust torisejate ravimisel*. [rahvahaal.delfi.ee]

(299) *Ära muretse..haiglas oled **arstide hoolitseva käe all** ja räägi kinldasti ära, et pole ümber keerand jne*. [www.nupsu.ee]

Nagu öeldud, viitab ühendi kasutamine tervikliku tähendusega üksusena leksikaliseerumisele. Liitpostpositsioonide tekkimine on käsitletav aga grammatiseerumisjuhuna, kuna protsessi lõpptulemus on grammatiline üksus – postpositsioon. Järgnevalt tutvustatakse liitpostpositsioonide tuvastamise kriteeriume, mis viitavad reanalüüsile – mehhanismile, mida tihti seostatakse grammatisatsiooniprotsessiga.

Desemantiseerumine. Uuritud kaassõnade semantiline analüüs nätab, et leksikaliseerunud üksused väljendavad abstraktsemaid tähendusi kui vabalt

kombineeruvad üksused. Selline tähendusmuutus on kooskõlas liitüksuste ümbertõlgendamisega liitseteks funktsioonisõnadeks, mis väljendavad funktsionaalseid kategooriad, nagu RUUM, AEG, VIIS jne, ning on kaugenenud oma kompositsioonilistest tähendustest. Selliseid tähendusmuutusi ei saa käsitada desemantiseerumisena, sest sellistes kasutustes kustub kehaosanimetuse referentsiaalne tähendus ning see asendub uue, grammatilise(ma) tähendusega. Desemantiseerumist peetakse grammatiseerumise üheks põhiliseks parameetriks (Heine, Kuteva 2002; 2007). Näiteks ühend *käe all* väljendab liitkaassõnana funktsiooni KONTROLL, mis võib avalduda nii füüsilise kui ka mentaalse kontrollina. Uuritavate fraaside desemantiseerunud kasutused on mõistagi seotud kompositsioonilise tähendusega ning paljud neist viitavad samasuunalisele arengule, mida on täheldatud ka teises keeltes ning mis on kooskõlas grammatiseerumise universaalsete arenguteedega. Selline on näiteks *selja taga* areng kehaosanimetusega seotud kaassõnafraasist liitüksuseks, mis väljendab tähendusi '(ruumiliselt) möödas' ja '(ajaliselt) möödas'. Mõnel juhul väljendavad aga ka liitüksused üsna spetsiifilisi tähendusi (nt *kaela peal* tähenduses 'koormaks').

Ekstensioon. Liitstruktuuri kuuluvad üksused saavad esineda kontekstis, mis pole lihtstruktuuri puhul võimalik või on väga harv. Keeleüksuse levimist uutesse kontekstidesse e ekstensiooni peetakse grammatisatsiooni üheks olemuslikuks parameetriks (Heine, Kuteva 2002; Heine 2003, Heine, Kuteva 2007). Ekstensiooni vaadeldi uuritavate fraaside koosesinemisel kollektiivile ja abstraktsetele entiteedile viitava noomeniga (*erakonna käe all*), kuna kollektiivne noomen on lihtstruktuuri puhul sobimatu (*erakonna käsi*). Fraasile eelneva, kuid samasse fraasistruktuuri kuuluva noomeni semantilist klassi vaadeldi ka eraldi lihtüksustes ja liitüksustes. Peaaegu kõik uuritavad fraasid (v.a *jalge all*) esinevad liitüksusena. Mitte ühegi uuritud fraasi puhul ei ole kollektiivsed ja abstraktsed laiendid eriti levinud, kuid proportsionaalselt kõige enam leidub neid fraasidega *selja taga* (9%) ja *kaela peal* (10%). Siiski on näha, et kollektiivsed ja abstraktsele entiteedile viitavad üksused on levinud vaid liitüksusena tõlgendatavate näidete hulgast. Siin on erandiks vaid *selja taga* mõned kasutusjuhud, kus kollektiivsed (ent mitte abstraktsed) noomenid moodustasid 1% (20 näidet 2178st) vabalt kombineeruva üksusena analüüsitud näidete hulgas. Tegelikult võiks selliseid juhte tõlgendada ka ruumitähendusega liitsete funktsioonisõnadena, kuid posterioossele regioonile viitava liitse funktsioonisõna ja vabalt kombineeruva kaassõnafraasi vahel ei ole võimalik ühemõtteliselt vahet teha. Ekstensioon viitab siin esimesele tõlgendusele, sest see näitab, et fraasi tähendus on kehaosanimetusest eemaldunud.

Olenemata sellest on kollektiivsetel kasutusjuhtudel selge seos inimesele viitava noomeniga, sest tavaliselt tähistavad need inimrühma, harvem institutsiooni. Abstraktsed tähendused on nii harvad, et neid ei ole enamasti (veel) võimalik üldistavalt analüüsida. Kokkuvõttes võib öelda, et kollektiivile (või abstraktsele entiteedile) viitavad noomenid esinevad põhiliselt vaid liitüksusena tõlgendatavatel juhtudel ning on sellisena käsitletavad ekstensiooni

indikaatoritena. Olenemata sellest, et sellised kasutusjuhud ei ole väga levinud, viitab nende olemasolu sellele, et olles minetamas seost oma kompositsioonilise tähendusega, hakkavad liitüksused levima uutesse kontekstidesse. Seda saab pidada grammatiseerumisele iseloomulikuks jooneks.

Kehaosanimetust sisaldavatele fraasidele eelneva noomeni semantilise klassi kaudu on võimalik uurida ka liitüksuste tekke võimalikke arenguradu. Vaadeldes noomeni semantilist klassi eraldi liht- ja liitüksuste puhul, selgitati välja nende distributsioon kummagi struktuuriüksuse hulgas. Eristati klassid, mis esinevad vaid lihtüksusega; klassid, mis esinevad vaid liitüksusega; ja klassid, mis esinevad mõlema struktuuriüksusega. Peaaegu kõigi fraaside puhul (v.a *külje all*) tuleb välja, et ühisosaks on inimesele viitav noomen, mis on ka kõige tõenäolisem muutuse kontekst. Nende fraaside puhul, mille esikomponent ei ole produktiivselt kasutusel objekti osana (*käe all*, *selja taga*, *käe kõrval*, *jalge all*),[75] on selgelt täheldatav elusolendist (enamasti inimesest) lähtuv tähendusmuutus. See osutab, et nihe lihtstruktuurist liitstruktuuriks on tekkinud just inimesega seotud kontekstis. Selline tulemus on üsna ootuspärane nende funktsionaalsete kategooriate kontekstis, mida liitüksused esindavad – näiteks (mentaalne) KONTROLL (*käe all*), SALAJA (*selja taga*), TOETUS (*selja taga*), VARJATUS (*selja taga*), KOOREM (*kaela peal*) – on üsna selgelt inimesega seotud mõisted. Siiski võiks ruumi- ja ajafunktsioonisõnade puhul (ruumiline ja ajaline *selja taga*) oodata ka arengut, mis järgib varasemates uurimustes välja pakutud ahelat, mille kohaselt kehaosanimetuste grammatiseerumine ruumitähendusega funktsioonisõnadeks saab alguse kehaosale viitavast nimisõnast, mis laieneb kõigepealt objekti osa tähistavaks nimisõnaks ning seejärel funktsioonisõnaks (Svorou 1994; Heine 1997). Et aga *selja taga* esineb koos objektile viitava noomeniga väga harva, ei ole selline areng kuigi tõenäoline. Pealegi esinevad ruumilis-ajalised tähendused enamasti koos inimesele viitava või kollektiivse laiendiga.

Uuritavate fraaside seas oli potentsiaalselt kaks üksust, mille esikomponenti kasutatakse produktiivselt objekti osale viitamisel – *külje all* ja *kaela peal*. Materjali analüüsist selgus, et fraasi *kaela peal* kasutatakse siiski peamiselt inimesega seotud kontekstides (seda nii liht- kui ka liitstruktuuri puhul). Fraasi *külje all* aga kasutatakse nii inimesele kui ka objektile viitavana ning selle üksuse analüüsist ei selgunud, kas nihe lihtstruktuurist liitstruktuuriks on tekkinud inimese või elutu objekti kontekstis. Nii inimesele kui ka objektile viitavaid noomeneid leidus nii vabalt kombineeruvate üksuste kui ka liitpostpositsioonide seas. Samas kollektiivile ja regioonile viitavaid noomeneid leidus vaid liitüksuste seas. Need semantilised klassid on loogiliselt seostatavad vastavalt inimreferendi ja objektile viitava referendiga. Raske on öelda, kummas kontekstis tähendusmuutus varem toimus. On tõenäoline, need on paralleelselt toimunud arengud.

---

[75] Sõna *jalg* on küll objekti osana kasutusel, kuid lühike mitmuse omastava vorm *jalge* viitab selgelt vaid kehaosale, mitte objekti osale.

Dekategoriseerumine. Tänapäeva keele materjalis on näha, et uuritavat fraasi võivad laiendada nii ainsuslik kui ka mitmuslik noomen. Kuna mitmusliku noomeni kasutamist võib pidada vabalt kombineeruvate üksuste puhul semantiliselt sobimatuks, (*õpetajate käsi* > *[[õpetajate käe] all]*), sest semantiline ühildumatus kehaosanimetuse ja sellele eelneva noomeni vahel viitab sellele, et fraasisuhted on ümber tõlgendatud ja et eelnev noomen ei seostu enam otseselt kehaosanimetusega, vaid laiendab tervet fraasi. Siiski näitab materjal enamiku (v.a *käe kõrval*) uuritavate fraaside puhul, et ühildumatust avaldub nii lihtstruktuuri kui ka liitstruktuuri puhul, kuid tavalisem on ühildumatus liitkaassõnade puhul. Samas on võimalik, et ühildumatus ei viita tingimata dekategoriseerumisele, vaid tuleneb kehaosanimede nn ainsuslikukst mitmusest (Alvre 1968). Analüüs näitas, et uuritavate kehaosanimedega seotud kaassõnafraaside esikomponent esineb üldisemalt mitmuses harva. On võimalik, et see tuleneb asjaolust, et kujundlikku väljendina eeslistab fraas ainsuslikku malli (Õim ja Õim (2015)).

2. Missugune on sageduse roll liitsete funktsioonisõnade arengus? Kuivõrd kokkukuuluvaks ja produktiivseks võib uuritavaid fraase pidada?

Tegelikule keelekasutusele tuginemine võimaldab arvesse võtta ka uuritavate üksuste sagedust. Lisaks absoluutsetele ja suhtelistele sagedustele rakendatakse uurimuses ka keerukamat statistilist meetodit, millega mõõdetakse sõnadevahelise seose tugevust. Selleks kasutatakse log-tõepära funktsiooni, mille rakendamisel on kaks eesmärki. Fraasi osiste seoste tugevus näitab seda, kuivõrd on fraas kinnistunud, st see võimaldab mõõta, kuivõrd kokkukuuluvad on uuritavad fraasid. Teiselt pool on seda meetodit kasutatud fraasi kui terviku ja selle vahetus lauseümbruses paiknevate üksuste – verbi ja eelneva noomeni – vahelise seose tugevuse mõõtmiseks. See näitab, kuivõrd produktiivselt liitüksusi kasutatakse, st kuivõrd sõltuv on liitüksuste kasutus ümbritsevast lausekontekstist või kuivõrd erinevates kontekstides liitseid üksusi kasutada saab.

Analüüs näitab, et uuritavate fraaside komponendid on üsna tihedalt seotud – seda nii võrreldes fraasidega, mille esikomponent on mitmuslik (*selja taga* vs. *selgade taga*), kui ka võrreldes vabalt kombineeritavate fraasidega, mis olid moodustatud uuritavate kehaosanimetuste ja kaassõnade vabal kombineerimisel, mille ainsaks kriteeriumiks oli see, et fraas ei ole absurdse tähendusega ning väljendab rohkem või vähem tõenäolist suhet (*nina kõrval*). See, et uuritavad fraasid said selles analüüsis mitu korda kõrgemad skoorid, viitab sellele, et tegemist on kokkukuuluvate üksustega. On selge, et kõikide fraaside puhul ei viita skoor sama tugevale seosele. Siiski võib öelda, et üldjoontes vastab sõnadevahelise seose tugevuse skoori alusel tehtud järjestus absoluutse sageduse alusel koostatud järjestusele. See tähendab, et vaieldamatult tugevaim seos on fraasi *selja taga* osiste vahel, sellele järgnevad *käe all*, *külje all* ja *jalge all* keskmise tugevusega ja kõige nõrgemalt on seotud vähem sagedad *kaela peal* ja *käe*

kõrval. Sellest hoolimata võib kõiki uuritavaid fraase pidada kokkukuuluvateks üksusteks.

Kui vaadelda liitpostpositsioonide produktiivsust liitpostpositsioonina esineva fraasi ja sellele eelneva noomeni ning sellega koos esineva verbi seose tugevuse alusel, ilmneb, et kuigi kõikidel uuritavatel fraasidel on ka postpositsioonilises kasutuses oma kindlad kollokaadid, on sagedasemate fraaside postpositsiooniline kasutus üsna mitmekesine. Näiteid, kus fraas esineb koos tugevate kollokaatidega, võib pidada enamasti prototüüpsemateks kasutusjuhudeks. Näiteks *juhendaja käe all*, *õpetaja käe all* jms on liitkaassõna *käe all* tüüpilised kasutusjuhud, kuid *instruktori* või *spetsialisti käe all* on vähem tüüpilised. Siiski selgus materjali analüüsil, et enamasti ei moodusta uuritavad fraasid koos tugevamate kollokaatidega idiomaatilisi väljendeid (v.a ehk (*nagu*) *vanajumala selja taga*), kuid sellest hoolimata ei saa selliseid kinnistunud üksusi pidada produktiivse kasutuse näideteks. Nõrgemate kollokaatidega koosesinemine näitab aga, et liitpostpositsioon on võimeline moodustama kaassõnafraase erinevate laienditega, mis omakorda näitab selle produktiivsust. Suurt produktiivsust on tihti seostatud grammatisatsiooniga. Analüüsis selgus, et produktiivsemateks võib uuritavate fraaside seast pidada liitpostpositsioone *käe all*, *külje all* ning vähem produktsiivsemateks fraase *kaela peal* ja eriti fraasi *jalge all*. Teisi fraase, sh mitut funktsiooni kandvat *selja taga*, võib pidada kesmiselt produktiivseks. Üldiselt võib öelda, et sagedamate ja üldisema tähendusega fraaside kasutus on mitmekesisem, erandiks on vaid uuritavate fraaside seas keskmise sagedusega *jalge all*, mida võib ka teiste parameetrite alusel pidada siinses töös uuritud fraasidest kõige vähem grammatiseerunuks.

6.4.2. Diakroonilise analüüsi tulemused

1. Milliseid tõendeid leidub uuritavate liitsete funktsioonisõnade diakroonilise arengu kohta? Mis ajast pärinevad uuritavate fraaside esimesed kasutused liitüksustena? Kas aja jooksul on märgata liitüksuste osakaalu suurenemist? Kas uuritavatel perioodidel ilmneb märke grammatisatsioonile viitavate faktorite levimisest?

Kuna enamiku uuritavate fraaside kohta oli materjali üsna vähe, tuleb diakroonilise uurimuse tulemustesse suhtuda ettevaatusega, sest vähene andmehulk suurendab juhuslikkuse rolli ning olemasolev materjal ei pruugi näidata fraaside tegelikku arengut. Sellest hoolimata võimaldab diakrooniline andmestik anda siinsele uurimusele diakroonilise mõõtme, võimaldades vaadelda liitüksuste arengut perioodide lõikes: 17.–19. sajandi kirjakeeles, 20. sajandi esimesel poolel ja 20. sajandi teisel poolel. Diakroonilises uurimuses vaadeldi mh samu aspekte mida sünkroonilises materjalis, s.o üldist sagedust, liht- ja liitstruktuuride suhtelist sagedust, ekstensiooni ja dekategorisatsiooni. Analüüs näitas, et fraaside üldine sagedus kasvab uuritavate perioodide jooksul (*käe all*, *selja taga*, *külje all*, *käe kõrval*, *jalge all*). Fraasi *kaela peal* kasutuses ilmnes aga

20. sajandi esimesel poolel sageduse järsk langus, mille võimaliku seletusena võib näha liitüksuse väljatõrjumist samatüvelise ja sama funktsiooniga liht-tüvelise funktsioonisõna poolt (*kaelas*). Teiste seas võis vaadeldud perioodidel täheldada liitüksuste suhtelise sageduse suurenemist (v.a fraasi *käe kõrval* puhul, kus see püsis stabiilselt suurena kõikidel vaadeldavatel perioodidel). Ekstensiooni esines diakroonilises materjalis väga vähe. On võimalik, et ka tänapäeval harva esinevate kollektiivsete laiendite kasutusjuhud olid varem olemas, kuid ei sattunud korpusesse. Samal ajal on ka võimalik, et selliseid juhte ei esinenud standardkeeles, mida 20. sajandi kirjakeele korpus sisaldab. Tegemist võib olla ka hilisema arenguga, mis alles tänapäeval vähehaaval levib. Dekategoriseerumine oli diakroonilises materjalis rohkem levinud, kuid selle kriteeriumi puhul oli fraasidevaheline varieerumine suurim. Mõnel juhul ilmub ühildumatus materjali alles 20. sajandi teisel poolel (*käe all*, *käe kõrval*), mõnel juhul kõigub selle sagedus vaadeldud perioodidel (*selja taga*). Nagu tänapäeva keeleski, esinevad mitmuslikud vormid enamasti nii liht- kui ka liitstruktuuriga, v.a fraas *käe all*, mille puhul mitmuslikud vormid esinevad vaid liitüksustega.

2. Kas sellised üksused, mis esinevad tänapäeva keeles nii adverbi kui ka postpositsioonina, on kujunenud adverbialiseerumise ja seejärel kaas-sõnastumise teel (Habicht, Penjam 2007) või on teatud juhtudel võimalik ka vastupidine arengutee?

Adverbi ja postpositsioonide distributsioon oleneb eeskätt konkreetsest fraasist ja selle tähendusest. Sealjuures näitab materjal mõlemasuunalisi tendentse: adverbiaalse kasutuse suurenemist (nt *käe kõrval*) ja ka postpositsioonilise kasutuse osakaalu suurenemist (nt *selja taga* TOETUSE funktsioonis), mõne fraasi puhul on jagunemine perioodide lõikes stabiilne (nt *külje all*). Mis puutub adverbilise kasutuse eelnemisse postpositsioonilisele kasutusele, siis võib diakroonilise analüüsi tulemusel üsna kindlalt väita, et kuigi adverbiline vaheaste on teatud juhtudel võimalik ja loogiline, ei ole see eeldus liitse funkt-sioonisõna tekkimiseks. Näiteks fraasi *selja taga* funktsiooni VARJATUS puhul näitab materjal veenvalt, et liitne üksus on tekkinud just postpositsioonina ning adverbilised kasutused on tekkinud alles hiljem (või on nii marginaalsed, et neid ei esine mujal kui tänapäeva internetikeele korpuses). Sellisele arengule viitab (vähema kindlusega) ka fraaside *käe all* ja *kaela peal* diakrooniline analüüs.

3. Missugust rolli mängib leksikaliseerumine liitpostpositsioonide tekke-protsessis?

Siinse käsitluse järgi ei ole kaassõnafraaside leksikaliseerumine seotud otseselt nende adverbistumisega, vaid üldisemalt uue liitüksuse tekkimisega ja selle vastuvõtmisega (mentaalsesse) leksikoni (vt Brinton, Traugott 2005). Leksikali-satsioon on siin käsitletud grammatisatsiooniprotsessis olulisel kohal, kuna

uurimisobjekt on mitmes mõttes tavatu grammatisatsioonijuht. Tegemist ei ole mitte üksiku sõnavormi grammatiseerumisega, vaid protsessiga, kus süntaktiline fraas (kaassõnaühend) siseneb tervikuna grammatiseerumisprotsessi. Selle eelduseks on aga fraasi eelnev leksikaliseerumine, terviku tõlgenduse saamine.

Leksikalisatsiooni vaheaste grammatisatsiooniprotsessis lisab grammatisatsiooniprotsessile iseäralikke jooni. See võimaldab suhteliselt abstraktsete üksuste üsna järsku tekkimist keelde. Sellega ei peeta silmas terviktähenduste järsku teket või selle protsessi produktiivsust, sest ka leksikalisatsiooni peetakse järkjärguliseks protsessiks (Brinton, Traugott 2005). Leksikalisatsioon võimaldab selles grammatisatsiooniprotsessis tekkida üsna abstraktsetel tähendustel (nagu näiteks mentaalne (või) füüsiline KONTROLL, kellegi TAGASELJA tegutsemine või kellelegi (vaimseks) KOORMAKS olemine) ühekorraga. Tüüpilises grammatisatsiooniprotsessis võiks arvata, et seesuguste tähenduste kujunemine võtab pikka aega ja selleks peab grammatiseeruv üksus läbima palju etappe, kuid leksikalisatsioon võimaldab üksusel läbida grammatisatsiooniprotsessi alguses suurema semantilise nihke, millele hakkavad alles hiljem lisanduma tüüpilised grammatisatsiooni tunnused, nagu ekstensioon või dekategoriseerumine ja produktiivsuse suurenemine.

4. Kas uuritavad liitpostpositsioonid on tekkinud loomuliku keelemuutuse käigus või on tegemist keelekontakti mõjul toimunud keelemuutusega?

Enamiku fraaside puhul puuduvad diakroonilises materjalis andmed võõrmõju kohta. Kuid kahe fraasi puhul – *käe all* ja *kaela peal* – viitab materjal sellele, et nende arengut on võinud mõjutada saksa keel. Fraasis *käe all* viitab sellele mitme näite saksakeelne paralleeltekst, *kaela peal* puhul leidus Wiedemanni sõnastikus (1973) [1893] sõnasõnaline saksakeelne vaste. Seega näitab materjal, et *käe all* ja *kaela peal* on ilmselt leksikaliseerunud saksakeelsete eeskujude mõjul. Kuid kuna need saksakeelsed väljendid ei ole grammatilised üksused, siis saab seda protsessi käsitada pigem kui tõlkelist laenamist, mitte keelekontaktist mõjutatud grammatiseerumist (*contact induced grammaticalization*), Heine, Kuteva 2005). Asjaolu, et teiste fraaside puhul paralleeltekst puudub, ei välista iseenesest võimalikku saksa mõju, sest saksa paralleeltekste ei sisaldu kõigis allikates. Siiski on enamik teisi uuritavaid fraase Wiedemanni sõnaraamatus vastavate kehaosanimetuste artiklite all välja toodud ja nende saksa vasted ei ole sõnasõnalised. See osutab, et 19. sajandi teiseks pooleks olid need fraasid küll leksikaliseerunud, kuid otsene saksa eeskuju neil ilmselt puudus.

6.5. Kokkuvõtteks

Uurimuse tulemused näitasid, et kehaosanimetustega seotud kaassõnafraasid reanalüüsitakse liitpostpositsiooniks peamiselt inimesega seotud kontekstides. Selles protsessis on oluline roll ühendi tähenduse muutumisel, st fraasi

tervikuna tõlgendamisel, mis on olulisim kriteerium liitüksuste tekkimisel. Leksikaliseerumine võimaldab üksusel liikuda grammatisatsiooniprotsessis edasi, kuna see annab üksusele algvormiga võrreldes abstraktsema tähenduse. Selle toel toimubki üksuse reanalüüs, milles varasem vabalt kombineeruv fraas tõlgendatakse ümber liitseks terviküksuseks, mis funktsioneerib postpositsioonina. Reanalüüs on aga nähtamatu mehhanism ning selle toimumist on võimalik vaadelda vaid selle aktualiseerumise järel. Siinse nähtuse puhul viitavad reanalüüsi toimumisele ekstensioon ja dekategoriseerumine. Need nähtused ei ole ka tänapäeva keeles uuritavate üksuste puhul eriti laialdaselt levinud, ent nende olemasolu ning kalduvus realiseeruda kontekstides, kus fraasid kannavad terviktähendust, viitab selgelt üksuste grammatiseerumisele.

Ekstensioon on analüüsi põhjal selgemalt seotud liitüksustega, kuid dekategoriseerumist on mõnel määral märgata ka lihtüksuste puhul. See tähendab, et ainult liitpostpositsioonile võib eelneda kollektiivset referenti tähistav laiend. Ühildumatust esineb mõlema struktuuriga seoses, kuid tavalisem on see liitstruktuuri puhul. Osalt võib selliseid näiteid tõlgendada ka liht- ja liitstruktuuri vahepealsete juhtudena. Selline käsitlus on kooskõlas grammatisatsiooni üldpõhimõtetega, mille kohaselt saab grammatiseerumine olla vaid aeglane järkjärguliselt kulgev protsess, ja mõnede käsitlustega reanalüüsist, mille kohaselt ei ole ka reanalüüsis järske üleminekuid, vaid nihe ühest struktuurist teise võib sisaldada nn vahestaadiume, kus grammatiseeruval keeleüksusel võib olla tunnuseid, mis viitavad selle kuulumisele korraga nii lähte- kui ka sihtvormi. Diakroonilisest analüüsist selgus, et adverbistumise etapp ei pruugi liitpostpositsioonide tekkimisele tingimata eelneda: võimalik on ka vastupidine kujunemistee. Samuti osutab diakrooniline materjal, et osa uuritavaid liitpostpositsioone on tekkinud saksa keele mõjul, ent mitte grammatilise kopeerimise kaudu, vaid tõlkelaenulisel teel.

ABBREVIATIONS

List of abbreviations used in the dissertation

ADV	adverb	PL	plural
BP	body part	PostP	postposition
LM	landmark	SG	singular
LOC	locative function word	CELL	Corpus of Literary Estonian
TR	trajector		
OP	object part	COLE	Corpus of Old Literary Estonian
pmw	per million words		
PN	preceding (pro)noun		

List of abbreviations used in the glosses

1, 2, 3	person	INE	inessive
ABE	abessive	INF	infinitive
ABL	ablative	LAT	lative
ADE	adessive	LOC	locative
ALL	allative	NEG	negation
CL	clitic	NOM	nominative
COM	comitative	PART	particle
COMP	comparative	PL	plural
COND	conditional	PRS	present
CONNEG	connegative	PRT	partitive
ELA	elative	PST	past
ESS	essive	PTCP	participle
GEN	genitive	SEP	separative
GER	gerundium	SG	singular
ILL	illative	SUP	supine
IMP	imperative	TER	terminative
IMPS	impersonal	TRL	translative

REFERENCES

Aavik, Johannes 1912. *Mõned keele reeglid. Nõuded ja soovid* [*Some rules of language. Requirements and suggestions*]. Tartu: Noor Eesti Kirjastus.

Aavik, Johannes 1913. Keele kaunima kõlavuse poole [Towards the more beautiful sound of language]. *Eesti Kirjandus*. Tartu: Eesti Kirjastuse-Ühisus Postimees, 3–36.

Adler, Silvia 2008. French compound prepositions, prepositional locutions and pre-positional phrases in the scope of the absolute use. In Dennis Kurzon, Silvia Adler (eds.) *Adpositions: Pragmatic, Semantic and Syntactic Perspectives*. Typological Studies in Language vol. 74. John Benjamins Publishing Company, 17–36.

Ahn, Mikyung 2010. From 'favor' to 'cause': the English causal adposition *thanks to* and its Korean counterpart *tekpwuney*. *Language Sciences* 32, 579–587.

Alvre, Paul 1989. *Eesti kirja- ja murdekeele morfoloogiat* [*The morphology of standard Estonian and dialects*]. Eesti NSV Teaduste Akadeemia Emakeele Seltsi toimetised 23. Tallinn: Valgus.

Arriola, Jose Maria 2012. Processing Euskara complex postpositions in a rule based approach in order to improve shallow syntactic disambiguation. In *31st International Conference on Lexis and Grammar With the Thematic Session Adverbs and Adverbial Complements*, 19–22 September 2012, Nové Hrady (Czech Republic), 5–9.

Beckner, Clay, Bybee, Joan L. 2009. A usage-based account of constituency and re-analysis. In Nick C. Ellis, Diabe Larsen-Freeman (eds.) *Language as a Complex Adaptive System. Language Learning*. A Journal of Research in Language studies 59 (60th anniversary special issue). Wiley & Sons, 27–46.

Bouma, Gosse, Begoña Villada 2002. Corpus-based acquisition of collocational prepositional phrases. *Language and Computers* 45: 1, 23–37. http://www.let.rug.nl/~gosse/papers/bouma_villada.pdf (Accessed 25.05.2015)

Brinton, Laurel, J., Elizabeth Closs Traugott 2005. *Lexicalization and Language Change*. Cambridge: Cambridge University Press.

Bybee, Joan L. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins.

Bybee, Joan L. 2003. Mechanisms of change in grammaticization: the role of fre-quency. In Brian. D. Joseph, Richard D. Janda (eds.) *The handbook of historical linguistics*, 602–623.

Bybee, Joan L. 2006. From usage to grammar: the mind's response to repetition. *Language* 82: 4, 711–733.

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

Bybee, Joan L. 2010. *Language, Usage and Cognition*. Cambridge University Press.

Choi-Jonin, Injoo 2008. Particles and postpositions in Korean. In Dennis Kurzon, Silvia Adler (eds.) *Adpositions: Pragmatic, Semantic and Syntactic perspectives*. Typological Studies in Language vol. 74. John Benjamins Publishing Company, 133–170.

Church, Kenneth Ward, Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics* 16: 1, 22–29.

Crystal, David 2000. *Dictionary of Linguistics and Phonetics*, 4th edition, updated and enlarged. Oxford: Blackwell Publishers.

Crystal, David 2011. *Internet Linguistics: A Student Guide*. London/New York: Routledge.

de Ilarraza, Arantza Dıaz, Koldo Gojenola, Maite Oronoz 2008. Detecting erroneous uses of complex postpositions in an agglutinative language. *Proceedings of COLING-08: Companion Volume – Posters and Demonstrations*, Manchester, 31–34.

De Smedt, Hendrik 2012. The course of actualization. *Language* 88, 601–633.

Eckardt, Regine 2006. *Meaning Change in Grammaticalization: An Enquiry into Semantic Reanalysis*. New York: Oxford University Press.

Ehasalu Epp, Külli Habicht, Valve-Liivi Kingisepp, Jaak Peebo 1997. *Eesti keele vanimad tekstid ja sõnastik* [*The Oldest Texts of Estonian and the Dictionary*]. Tartu Ülikooli eesti keele õppetooli toimetised vol. 6. Tartu: Tartu University Press.

Erelt, Mati (Ed.) 2003. *Estonian Language*. Tallinn: Estonian Academy Publishers.

Erelt, Mati 2009. Typological overview of Estonian syntax. *STUF*, Akademie Verlag, 62: 1/2, 6–28.

Erelt, Mati 2013. *Eesti keele lauseõpetus. Sissejuhatus. Öeldis* [*Estonian Syntax. An Introduction. A Predicate*]. Tartu Ülikooli eesti keele osakonna preprindid 4 (Preprints of the Department of Estonian of the University of Tartu). Tartu: Tartu Ülikool.

Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael and Silvi Vare 1993. *Eesti keele grammatika II. Süntaks* [*The Grammar of Estonian II. Syntax*]. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.

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

Evert, Stefan 2005. *The Statistics of Word Cooccurrences: Word Pairs and Collocations*. Ph.D. Dissertation. Institut für maschinelle Sprachverarbeitung Universität Stuttgart.
http://elib.uni-stuttgart.de/opus/volltexte/2005/2371/pdf/Evert2005phd.pdf?q=15-collocations (Accessed 12.01.2016)

Evert, Stefan 2008. Corpora and collocations. In Anke Lüdeling, Merja Kytö (eds.) *Corpus Linguistics. An international handbook*. Berlin: Mouton de Gruyter, 1212–1248.

Gries, Stefan 2014. Frequency tables: Tests, effect sizes, and explorations. In Dylan Glynn, Justina Robinson (eds.) *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, 365–389.

Grünthal, Riho 2003. *Finnic Adpositions and Cases in Change*. Suomalais-ugrilaisen seuran toimituksia 244. Helsinki: The Finno-Ugrian Society.

Habicht, Külli 2000. Grammaticalization of adpositions in Old Literary Estonian. In Mati Erelt (ed.) *Estonian: Typological Studies IV*. Tartu Ülikooli eesti keele õppetooli toimetised vol. 14. Tartu, 19–58.

Habicht, Külli 2001a. *Eesti vanema kirjakeele leksikaalsest ja morfosüntaktilisest arengust ning Heinrich Stahli keele eripärast selle taustal* [*About the Lexical and Morphosyntactic Development of Older Written Estonian and the Background for the Peculiar Features in the Language of Heinrich Stahl*]. Ph.D. Dissertation. Tartu: Tartu University Press.

Habicht, Külli 2001b. On the genesis and loss of the adposition *rinnas* ('abreast, beside') in Literary Estonian. In Ilona Tragel (ed.). *Papers in Estonian Cognitive Linguistics*. Tartu: Tartu University Press, 71–89.

Habicht, Külli 2001c. Lexicalization and grammaticalization – opposite phenomena? About some lexicalized verb forms of Old Written Estonian. In Mati Erelt (ed.). *Estonian: Typological Studies V*. Tartu Ülikooli eesti keele õppetooli toimetised vol. 18, 26–76.

Habicht, Külli 2003. Vana kirjakeele uurimine [Studies of old written Estonian]. *Emakeele Seltsi aastaraamat* 48 (2002), 135–165.

Habicht, Külli, Pille Penjam 2007. Kaassõna keeleuurija ja -kasutaja käsituses [Adpositions as viewed by a linguist and by a language user]. *Emakeele Seltsi aastaraamat* 52 (2006), 51–68.

Habicht, Külli, Valve-Liivi Kingisepp, Urve Pirso, Külli Prillop 2000. *Georg Mülleri jutluste sõnastik* [*The Dictionary of George Müller's Sermons*]. Tartu Ülikooli eesti keele õppetooli toimetised vol. 12.

Harris, Alice C., Lyle Campbell 1995. *Historical Syntax in Cross-Linguistic Perspective*. Cambridge Studies in Linguistics vol. 74. Cambridge: Cambridge University Press.

Haspelmath, Martin 1998. Does grammaticalization need reanalysis? *Studies in Language* 22: 2, 315–351.

Haspelmath, Martin 2011a. The gradual coalesence into 'words' in grammaticalization. In Heiko Narrog and Bern Heine (eds.) *The Oxford Handbook of Grammaticalization*. New York/Oxford: Oxford University Press, 342–355.

Haspelmath, Martin 2011b. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45: 2, 31–80.

Heine, Bernd 1989. Adpositions in African linguistics. *Linguistique Africane* 2, 77–127.

Heine, Bernd 1997. *Cognitive Foundations of Grammar*. New York/Oxford: Oxford University Press.

Heine, Bernd 1999. Grammaticalization chains across languages: An example from Khoisan. In Spike Gildea (ed.). *Reconstructing Grammar. Comparative Linguistics and Grammaticalization*. Typological Studies in Language 43, 177–200.

Heine, Bernd 2003. Grammaticalization. In Brian D. Joseph, Richard D. Janda (eds.) *The Handbook of Historical Linguistics*, 575–601.

Heine, Bernd, Tania Kuteva 2002. *The World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.

Heine, Bernd, Tania Kuteva 2007. *The Genesis of Grammar. A Reconstruction.* Studies in the Evolution of Language. Oxford/New York: Oxford University Press.

Heine, Bernd, Ulrike Claudi, Frederike Hünnemeyer 1991. *Grammaticalization: A Conceptual Framework*. Chicago/London: University of Chicago Press.

Himmelmann 2004. Lexicalization and grammaticalization: Opposite or orthogonal. In Walter Bisang, Nikolaus Himmelamnn, Björn Wiemer (eds.) *What Makes Grammaticalization. A Look from its Fringes and its Components*. Trends in Linguistics 158. Mouton de Gruyter, Berlin/New York, 21–42.

Hoffmann, Sebastian 2005. *Grammaticalization and English Complex Prepositions: A Corpus-Based Study*. Routledge Advances in Corpus Linguistics. London: Routledge.

Hoffmann, Sebastian 2004. Are low-frequency complex prepositions grammaticalized? In Hans Lindquist, Christian Mair (eds.) *Corpus Approaches to Grammaticalization in English*. Studies in Corpus Linguistics 13, 171–210.

Hopper, Paul J. 1991. On some principles of grammaticization. In Elizabeth Closs Traugott, Bernd Heine (eds.) *Approaches to grammaticalization* vol. I. Amsterdam/

Philadelphia: John Benjamins Publishing Company. Typological Studies in Language 19: 1, 17–35

Hopper, Paul J., Elizabeth Closs Traugott 2003 [1993]. *Grammaticalization*, 2nd ed. Cambridge: Cambridge University Press.

Jakubíček, Miloš; Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý and Vít Suchomel 2013. The tenten corpus family. In *Proceedings of the 7th International Corpus Linguistics Conference CL 2013, United Kingdom, July 2013*, 125–127. http://www.sketchengine.co.uk/documentation/raw-attachment/wiki/SkE/Biblio/The_TenTen_Corpus_2013.docx (Accessed 30.05.2015)

Janda, Laura A. 2006. Cognitive Linguistics. *Glossos* vol 8. http://seelrc.org/glossos/issues/8/janda.pdf (Accessed 29.09.2010)

Jürine, Anni 2009. Kas **metsavahel* tähendab muud kui *metsa vahel*? Oma Keel 18, 38–43.

Jürine, Anni 2011. Vormierinevused tähenduserinevuse ilminguna: kaassõnaühendite grammatiseerumine konstruktsioonides. [Change in Form as a Manifestation of Change in Meaning: Estonian Adpositional Phrases] *Keel ja Kirjandus* 12, 899–916.

Jürine, Anni, Külli Habicht 2013. Kaassõnade tsükliline areng [The cycle of postpositions]. *Keel ja Kirjandus* 12, 736–749.

Kabak, Barış 2006. An obstacle to the morphologization of postpositions. *Studies in Language* 30: 1, 33–68.

Kallas, Jelena, Maria Tuulik 2015. etTenTen13, the first web corpus for Estonian. Presentation at International seminar *Web Archiving: Preserving the History of Data-Driven Society*, 28 January 2015, Tallinn. https://www.nlib.ee/en/23991/ (Accessed 30.05.2015).

Karelson, Rudolf 2005. Taas probleemidest sõnaliigi määramisel [On the problems of how to determine the word class, once again]. *Eesti Rakenduslingvistika Ühingu Aastaraamat 1* (2004), 53–70.

Kasik, Reet 2013. *Komplekssete sõnade struktuur.* [*The Structure of Complex Words*] Tartu Ülikooli eesti keele osakonna preprindid 3. Tartu: Tartu Ülikool.

Kikas, Kristel 2002. *Mida sisaldab Heinrich Stahli* Vocabula*?* [*What can be found in Heinrich Stahl's* Vocabula*?*] Tartu Ülikooli eesti keele õppetooli toimetised vol. 21. Tartu: Tartu University Press.

Kingisepp Valve-Liive, Kristel Ress, Kai Tafenau 2010. *Heinrich Gösekeni grammatika ja sõnastik 350* [*The Grammar and Dictionary of Heinrich Göseken 350*]. Tartu: Tartu University Press.

Klavan, Jane 2012. *Evidence in Linguistics: Corpus-Linguistic and Experimental Methods for Studying Grammatical Synonymy*. Ph.D. Dissertation. Tartu: Tartu University Press.

Krug, Manfred 1998. String frequency: A cognitive motivating factor in coalescence, language processing, and linguistic change. *Journal of English Linguistics* 26: 4, 286–320.

Krug, Manfred 2000. *Emerging English Modals: A Corpus-Based Study of Grammaticalization*. Berlin/New York: Mouton de Gruyter.

Lakoff, George, Mark Johnson 1999. *Philosophy in the Flesh. The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.

Langacker, Ronald W. 1977. Syntactic reanalysis. In Charles N Li (ed.). *Mechanisms of Syntactic Change*. Austin: University of Texas Press, 57–140.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Volume I: Theoretical prerequisites*. Standford: Standford University Press.

Langacker, Ronald W. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

Lehmann, Christian 1985. Grammaticalization: Synchronic variation and diachronic change. *Lingua e Stile* 20: 3, 303–318.

Lehmann, Christian 1991. Grammaticalization and related changes in contemporary German. In Elizabeth Closs Traugott, Bernd Heine (eds.) *Approaches to grammaticalization*. Volume II. Amsterdam/Philadelphia: John Benjamins Publishing Company. Typological Studies in Language 19, 493–535.

Lehmann, Christian 1998. German abstract prepositional phrases. In Iørn Korzen, Michael Herslund (eds.) Clause combining and text structure. Frederiksberg: Samfundslitteratur. Copenhagen Studies in Language 22, 87–106.

Lehmann, Christian 2002. New reflections on grammaticalization and lexicalization. In Ilse Wischer, Gabriele Diewald (eds.) *New Reflections on Grammaticalization*. Typological Studies in Language 49. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1–18.

Lehmann, Christian, Christel Scholz 1992. *Bildung von Adpositionen im Deutschen*. Erfurt: Seminar für Sprachwissenschaft der Universität Erfurt. ASSidUE, 6.

Lieber, Rochelle, Pavol Štekauer 2009. Introduction. In Rochelle Lieber, Pavol Štekauer (eds.) *The Oxford handbook of compounding*. Oxford/New York: Oxford University Press.

Mair, Christian 2004. Corpus linguistics and grammaticalization theory: Statistics, frequencies, and beyond. In Hans Lindquist, Christian Mair (eds.) *Corpus Approaches to Grammaticalization in English*. Studies in Corpus Linguistics 13, 121–150.

Mair, Christian 2011. Grammaticalization and corpus linguistics. In Bernd Heine, Heiko Narrog (eds.) The Oxford Handbook of Grammaticalization. Oxford/New York: Oxford University Press, 239–250.

McMichael, Andrew 2006. The A's and BE's of English prepositions. Syntax and semantics of prepositions. In Patrick Saint-Dizier (ed.). *Text, Speech and Language Technology* vol. 29. Springer, 43–56.

Metslang, Helle 2006. Grammatisatsiooniteooriast tänase pilguga [The theory of grammaticalization today]. In Ilona Tragel, Haldur Õim (eds.) *Teoreetiline keeleteadus Eestis II* [Theoretical Linguistics in Estonia]. Tartu Ülikooli üldkeeleteaduse osakonna toimetised vol. 7. Tartu: Tartu Ülikool, 176–195.

Mikone, Eve 2000. Idiomit ja kileliopillistuminen [Idioms and grammaticalization]. In Eeva Lähdemäki, Yvonne Bertills *Fennistica Fausta. In Honorem Mauno Koski Septuagenarii. Juhlakirja mauno Kosken 70-vuotispäivan kunniaksi 21.2.2000.* Åbo: Åbo Akademis tryckeri, 16–32.

Moirón, Begoña Villada, Gosse Bouma 2003. A corpus-based approach to the acquisition of collocational prepositional phrases. In *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002: Copenhagen, Denmark, August 13–17, 2002*, 153–158.

Moore, Kevin Ezra 2000. Metaphor, linguistic practice, and the temporal meanings of *gannaaw* 'Back' and *kanam* 'front' in Wolof. *Proceedings of the 25th Annual Meeting of the Berkeley Linguistics Society*, 225–237.

Ojutkangas, Krista 2001. *Ruumiinosannimien kieliopillistuminen suomessa ja virossa*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Õim, Asta 2000. *Fraseoloogiasõnaraamat* [*Phraseological Dictionary*]. 2nd, complemented edition. Tallinn: Eesti Keele Sihtasutus.
http://www.eki.ee/dict/frs/ [Accessed 04.01.2016]

Õim, Katre, Asta Õim 2015. Kehaosasõnade arvukategooriast [Number in some body-part nouns]. *Keel ja Kirjandus* 2, 88–99.

Palmeos, Paula 1985. *Eesti keele grammatika II. Kaassõna* [*The Grammar of Estonian II. The adposition*]. Tartu: TRÜ trükikoda.

Petré, Peter, Kristin Davidse, and Tinne Van Rompaey 2012. On ways of being on the way: Lexical, complex preposition and aspect marker uses. *International Journal of Corpus Linguistics* 17: 2, 229–258.

Rätsep, Huno 1978. *Eesti keele lihtlausete tüübid* [*Patterns of Estonian Simple Sentences*]. Tallinn: Valgus.

Rhee, Seongha 2004. *A Comparative Analysis of Grammaticalization of English and Korean Adpositions*. Studies in Modern Grammar 40, 209–233.

Rostila, Jouni 2004. Lexicalization as a way to grammaticalization. *Proceedings of the 20th Scandinavian Conference of Linguistics*, Helsinki, 7–9.
http://www.ling.helsinki.fi/kielitiede/20scl/Rostila.pdf (Accessed 04.01.2016)

Rostila, Jouni 2006. Storage as a way to grammaticalization. *Constructions* 1.
http://www.constructions-online.de/articles/345 (Accessed 13.10.2010)

Schwenter, Scott A., Elizabeth Closs Traugott 1995. The semantic and pragmatic development of substitutive complex prepositions in English. In Andreas H. Jucker (ed.). *Historical Pragmatics. Pragmatic Developments in the History of English*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 243–273.

Seppänen, Aimo, Rhonwen Bowen, Joe Trotta 1994. On the so-called complex propositions. *Studia Anglica Posnaniensia* 29, 3–29.

Sepper, Maria-Maren. 2006. *Indirektaal eesti 19. sajandi lõpu ja 20. sajandi aja- ja ilukirjanduskeeles* [*Indirect evidentiality in Estonian journalistic and fiction texts of the late 19th and the 20th centuries*]. Manuscript of Master's thesis defended at Tallinn University.

Sepper, Maria-Maren. 2007. Indirectal in literary Estonian. *Trames: Journal of the Humanities and Social Sciences* 11: 3, 299–323.

Shepherd, Tania Maria Granja 2014. Changing 'faces': A case study of complex prepositions in Brazilian Portuguese. In Tony Berber Sardinha, Telma de Lurdes São Bento Ferreira (eds.) *Working with Portuguese Corpora*. Bloomsbury Academic, 69–88.

Sigurd, Bengt 1993. The problems of multi-word prepositions and subjunctions. *Working Papers in Linguistics* vol. 40, 197–212.

Smith, Adam 2013. Complex prepositions and variation within the PNP construction. In Hilde Hasselgård, Jarle Ebeling, Signe Oksefjell Ebeling (eds.) *Corpus Perspectives on Patterns of Lexis*. Studies in Corpus Linguistics 57, 153–176.

Spencer, Andrew 2008. Does Hungarian have a case system? In Corbett, Greville G., Michael Noonan (eds.) *Case and Grammatical relations: Studies in Honour of Bernard Comrie*. Typological Studies in Language 81. Amsterdam: John Benjamins Publishing Company, 35–56.

Suutari, Toni 2006. Body-part names and grammaticalization. In Maarja-Liisa Helasvuo, Lyle Campbell (eds.) *Grammar from the Human Perspective: Case, Space and Person in Finnish*. Current Issues in Linguistic Theory 277. John Benjamins Publishing Company.

Suzanne Kemmer and Michael Barlow 2000. Introduction: A usage-based conception of language. In Michael Barlow and Suzanne Kemmer (eds.) *Usage-Based Models of Language*. Stanford: Center for the Study of Language and Information, 7–28.

Svorou, Soteria 1994. *The Grammar of Space*. Typological Studies in Language 25. John Benjamins Publishing Company.

Tauli, Valter 1966. *Structural Tendencies in Uralic Languages*. London/The Hague: Mouton.

Tauli, Valter 1973. *Standard Estonian Grammar. Phonology, Morphology, Word-formation. Vol. 1*. Studia Uralica et Altaica Upsaliensia 8. Uppsala: Almqvist & Wiksell.

Traugott, Eliszabeth Closs 2003. Constructions in Grammaticalization. In Joseph, Brian D., Richard D. Janda (eds.) *The Handbook of Historical Linguistics*, 624–647.

Traugott, Elizabeth Closs, Scott Schwenter 1995. The semantic and pragmatic development of substitutive complex prepositions in English. In Andreas H. Jucker (ed.). *Historical Pragmatics*. Amsterdam: John Benjamins Publishing Company, 243–273.

Trawiński, Beata 2003. A new application for raising in HPSG: Complex prepositions. *Proceedings of FGVienna: The 8th Conference on Formal Grammar*, 163–173.

Uiboaed, Kristel 2010. Statistilised meetodid murdekorpuse ühendverbide tuvastamisel [Statistical methods for phrasal verb detection in Estonian dialects]. *Estonian Papers in Applied Linguistics* 6, 307–326.

van Pareren, Remco 2013. Body part terms as a semantic basis for grammaticalization: a Mordvin cases study into spatial reference and beyond. *Language Sciences* 36, 90–102.

Veismann, Ann 2009. *Eesti kaas- ja määrsõnade semantika võimalusi* [*Semantics of Estonian Adpositions and Adverbs*]. Dissertationes Linguisticae Universitatis Tartuensis 11. Tartu: Tartu University Press.

Veismann, Ann 2001. On the conceptualization of time in Estonian. In Ilona Tragel (ed.) *Papers in Estonian Cognitive Linguistics*. Tartu: Tartu University Press, 22−47.

Veski, Asta 1982. *Eesti keele grammatika. II. Kolmas vihik: Määrsõna* [*The Grammar of Estonian II. 3rd Part: The Adverb*]. Tartu Riiklik Ülikool, Tartu.

Villup, Astrid 1969. *Adverb eesti keeles* [*Adverb in Estonian*]. E. Vilde nim. Tallinna Pedagoogiline Instituut.

Zlatev, Jordan 2007. Spatial semantics. In Dirk Geeraerts, Hubert Cuyckens (eds.) *The Oxford Handbook of Cognitive Linguistics*, 318–350.

Annex 1. Components of possible body part related postpositional phrases

Body part terms		Spatial postpositions	
Jalg	'foot/leg'	*all*	'down/under'
Kael	'neck'	*ees*	'in front of'
Kand	'heel'	*hulgas*	'among'
Keha	'body'	*juures*	'near'
Kont	'bone'	*järel*	'after'
Kukal	'back of the neck'	*kannul*	'at heel'
Kulm	'eyebrow'	*keskel*	'in the middle of'
Kõht	'stomach/belly'	*killas*	'among'
Käsi	'hand/arm'	*kohal*	'above'
Külg	'side'	*kõrval*	'by'
Luu	'bone'	*küljes*	'on'
Lõug	'chin'	*ligidal*	'close'
Nahk	'skin'	*lähedal*	'near'
Nina	'nose'	*najal*	'leaned on'
Niuded	'loins'	*otsas*	'on top of'
Nägu	'face'	*peal*	'on'
Näpp	'finger'	*pool*	'towards'
Pea	'head'	*seas*	'among'
Piht	'waist'	*sees*	'in'
Puusad	'hips'	*seltsis*	'in the company of'
Põlv	'knee'	*taga*	'behind'
Põsk	'cheek'	*vahel*	'between'
Ranne	'wrist'	*veeres*	'beside'
Ribid	'ribs'	*ääres*	'beside'
Rind	'breast'		
Selg	'back'		
Silm	'eye'		
Suu	'mouth'		
Sõrm	'finger'		
Süda	'heart'		
Tald	'sole'		
Turi	'scruff'		
Varvas	'toe'		
Veri	'blood'		
Õlg	'shoulder'		

CURRICULUM VITAE

Name: Anni Jürine
Date of birth: 15.02.1985, Tallinn, Estonia
Citizenship: Estonian
Address: Institute of Estonian and General Linguistics,
               University of Tartu
               Ülikooli 18, 50090 Tartu, Eesti
E-mail: anni.jyrine@ut.ee

Education

2009–2016   University of Tartu, Ph.D. student in Estonian and Finno-Ugric Linguistics
2007–2009   University of Tartu, M.A. (Estonian and Finno-Ugric Linguistics), *cum laude*
2004–2007   University of Tartu, B.A. (Estonian and Finno-Ugric Linguistics)
2001–2003   Jakob Westholm Gymnasium
1997–2001   Tallinn Mustamäe Gymnasium
1991–1997   Tallinn Nõmme Secondary School

Professional employment

2012–…   S/A Innove; Estonian language proficiency exam assessor and developer
2015   University of Tartu, College of Foreign Languages and Cultures; lecturer
2014   University of Tartu, Centre for Academic Writing and Communication; assistant
2009–2011   The National Examination and Qualifications Centre; Estonian language proficiency exam assessor and developer

Awards

2014   Journal *Keel ja Kirjandus*, best paper in linguistics (2013) (co-author Külli Habicht)
2011   Societas Linguistica Europaea Annual Conference, Best Poster Presentation, 3rd prize

ELULOOKIRJELDUS

Nimi: Anni Jürine
Sünniaeg: 15. veebruar 1985, Tallinn
Kodakondsus: Eesti
Kontaktandmed: Tartu Ülikooli eesti ja üldkeeleteaduse instituut
Ülikooli 18, 50090 Tartu, Eesti
E-post: anni.jyrine@ut.ee

Haridustee
2009–2016 Tartu Ülikool, doktorantuur (eesti ja soome-ugri keeleteadus)
2007–2009 Tartu Ülikool, MA (eesti ja soome-ugri keeleteadus),
cum laude
2004–2007 Tartu Ülikool, BA (eesti ja soome-ugri keeleteadus)
2001–2003 Jakob Westholmi Gümnaasium
1997–2001 Tallinna Mustamäe Gümnaasium
1991–1997 Tallinna Nõmme Põhikool

Teenistuskäik
2012–… S/A Innove; eesti keele tasemeeksami hindaja ja ülesannete
koostaja
2015 Tartu Ülikool, maailma keelte ja kultuuride kolledž; lektor
2014 Tartu Ülikool, akadeemilise väljendusoskuse keskus; assistent
2009–2011 Riiklik eksami- ja kvalifikatsioonikeskus; eesti keele
tasemeeksami hindaja ja ülesannete koostaja

Tunnustused
2014 Ajakirja Keel ja Kirjandus aastapreemia (2013)
parima keeleteadusliku artikli eest (kaasautor Külli Habicht)
2011 Societas Linguistica Europaea aastakonverentsi 3. auhind
posterettekannete sessioonil

DISSERTATIONES PHILOLOGIAE ESTONICAE
UNIVERSITATIS TARTUENSIS

1. **Ülle Viks.** Eesti keele klassifikatoorne morfoloogia. Tartu, 1994.
2. **Helmi Neetar.** Deverbaalne nominaaltuletus eesti murretes. Tartu, 1994.
3. **Ülo Valk.** Eesti rahvausu kuradi-kujutelm. Tartu, 1994.
4. **Arvo Eek.** Studies on quantity and stress in Estonian. Tartu, 1994.
5. **Reet Kasik.** Verbid ja verbaalsubstantiivid tänapäeva eesti keeles. Tartu, 1994.
6. **Silvi Vare.** Nimi- ja omadussõnatuletus tänapäeva eesti kirjakeeles. Tartu, 1994.
7. **Heiki-Jaan Kaalep.** Eesti keele ressursside loomine ja kasutamine keeletehnoloogilises arendustöös. Tartu, 1998.
8. **Renate Pajusalu.** Deiktikud eesti keeles. Tartu, 1999.
9. **Vilja Oja.** Linguistic studies of Estonian colour terminology. Tartu, 2001.
10. **Külli Habicht.** Eesti vanema kirjakeele leksikaalsest ja morfosüntaktilisest arengust ning Heinrich Stahli keele eripärast selle taustal. Tartu, 2001.
11. **Pire Teras.** Lõunaeesti vokaalisüsteem: Võru pikkade vokaalide kvaliteedi muutumine. Tartu, 2003.
12. **Merike Parve**. Välted lõunaeesti murretes. Tartu, 2003.
13. **Toomas Help.** Sõnakeskne keelemudel: Eesti regulaarne ja irregulaarne verb. Tartu, 2004.
14. **Heli Laanekask.** Eesti kirjakeele kujunemine ja kujundamine 16.–19. sajandil. Tartu, 2004.
15. **Peeter Päll.** Võõrnimed eestikeelses tekstis. Tartu, 2005.
16. **Liina Lindström.** Finiitverbi asend lauses. Sõnajärg ja seda mõjutavad tegurid suulises eesti keeles. Tartu, 2005.
17. **Kadri Muischnek.** Verbi ja noomeni püsiühendid eesti keeles. Tartu, 2006.
18. **Kanni Labi.** Eesti regilaulude verbisemantika. Tartu, 2006.
19. **Raili Pool.** Eesti keele teise keelena omandamise seaduspärasusi täis- ja osasihitise näitel. Tartu, 2007.
20. **Sulev Iva.** Võru kirjakeele sõnamuutmissüsteem. Tartu, 2007.
21. **Arvi Tavast.** The translator is human too: a case for instrumentalism in multilingual specialised communication. Tartu, 2008.
22. **Evar Saar.** Võrumaa kohanimede analüüs enamlevinud nimeosade põhjal ja traditsioonilise kogukonna nimesüsteem. Tartu, 2008.
23. **Pille Penjam.** Eesti kirjakeele *da-* ja *ma*-infinitiiviga konstruktsioonid. Tartu, 2008.
24. **Kristiina Praakli**. Esimese põlvkonna Soome eestlaste kakskeelne keelekasutus ja koodikopeerimine. Tartu, 2009.
25. **Mari Mets.** Suhtlusvõrgustikud reaalajas: võru kõnekeele varieerumine kahes Võrumaa külas. Tartu, 2010.

26. **Karen Kuldnokk.** Militaarne retoorika. Argumentatsioon ja keeleline mõjutamine Eesti kaitsepoliitilises diskursuses. Tartu, 2011.
27. **Kai Tafenau**. Uue Testamendi tõlkimisest Rootsi ajal: käsikirjad, tõlkijad ja eesti kirjakeel. Tartu, 2011.
28. **Külli Prillop.** Optimaalsusteoreetiline käsitlus eesti keele fonoloogilisest kujunemisest. Tartu, 2011, 261 lk.
29. **Pärtel Lippus.** The acoustic features and perception of the Estonian quantity system, Tartu, 2011, 146 p.
30. **Lya Meister.** Eesti vokaali- ja kestuskategooriad vene emakeelega keelejuhtide tajus ja häälduses. Eksperimentaalfoneetiline uurimus. Tartu, 2011, 145 lk.
31. **Kersti Lepajõe.** Kirjand kui tekstiliik. Riigieksamikirjandite tekstuaalsed, retoorilised ja diskursiivsed omadused. Tartu, 2011, 141 lk.
32. **Tiit Hennoste.** Grammatiliste vormide seoseid suhtlustegevustega eestikeelses suulises vestluses. Tartu, 2013, 236 lk.
33. **Helena Metslang.** Grammatical relations in Estonian: subject, object and beyond. Tartu, 2013, 443 p.
34. **Kristel Uiboaed.** Verbiühendid eesti murretes. Tartu, 2013, 227 lk.
35. **Helen Plado.** Kausaalsuhete adverbiaallaused eesti keeles. Tartu, 2013, 244 lk.
36. **Annika Küngas.** Pragmaatiliste markerite kujunemine ja funktsioonid eesti keeles *lt*-sõnade näitel. Tartu, 2014, 200 lk.
37. **Maarika Teral.** Arvutipõhine eesti keele õpe: vahendid ja hinnangud nende efektiivsusele Tartu ülikooli keelekursuste näitel. Tartu, 2015, 175 lk.