

Adpositions in Estonian Computational Syntax

Kadri Muischnek

Department of General Linguistics
University of Tartu
J. Liivi 2, 50409, Tartu, Estonia
kadri.muischnek@ut.ee

Kaili Mürisep

Institute of Computer Science
University of Tartu
J. Liivi 2, 50409, Tartu, Estonia
kaili.muurisep@ut.ee

Tiina Puolakainen

Institute of Estonian Language
Roosikrantsi 6, 10119, Tallinn, Estonia
tiina.puolakainen@eki.ee

Abstract

In this paper we describe problems that arise in the analysis of adpositions in Estonian texts at the levels of manual corpus annotation, automatic morphological disambiguation and syntactic analysis. From our experiences we can conclude that the most important task is to determine precisely the exact boundaries of the class of adpositions, and then they do not cause any intractable problems at the various stages of automatic analysis.

1 Introduction

The Estonian language belongs to the Finnic group of the Finno-Ugric language family. Typologically Estonian is an agglutinating language but more fusional and analytic than the languages belonging to the northern branch of the Finnic languages. One can find a detailed description of the grammatical system of Estonian in (Erelt, 2003).

Despite the fact that there are 14 nominal cases in Estonian, the language makes use of adpositional phrases as well. Estonian has both pre- and postpositions. As adpositions are grammatical heads and determine the position and case form of the nominals participating in adpositional phrases, the correct analysis of these phrases plays an important role in the overall correctness of the syntactic analysis.

Adpositions turned out to be a complicated word class in the process of creating the morphological disambiguator and syntax analyzer for Estonian.

That is partly due to the fact that they do not form a really closed class in Estonian as they do, for example, in English. There is no clear-cut distinction between the case forms of some nouns and adpositions or between the non-finite forms of some verbs and adpositions. New adpositions (as well as adverbs) are constantly emerging from the inflectional forms of nominals and verbs as the result of a process of grammaticalization and these new emerging adpositions form ambiguous areas at the borders of word classes. So adposition lexicons tend to be insufficient. In addition to this, a great many of the word forms used as adpositions can also be adverbs and thus function as part of a particle verb.

In this paper we give a linguistic overview of the adpositions as a word class and as heads that govern the nouns in adpositional phrases. We also describe the problems that arise in the analysis of adpositions in Estonian texts at the levels of manual corpus annotation, automatic morphological disambiguation and syntactic analysis.

We began developing the morphological disambiguator and syntax analyzer of Estonian in the mid-nineties. We had the Corpus of Written Estonian (Hennoste et al., 1998) at our disposal. Also a morphological analyzer for Estonian had been developed (Kaalep, 1996), but it did not differentiate between pre- and postpositions. In parallel with the development of the formal grammar a lexicon of adpositions was compiled that was equipped with information about government.

We use the Constraint Grammar formalism (Karlsson et al., 1995) for the automatic analysis of Estonian.

2 Constraint Grammar

The main idea of Constraint Grammar (CG) is that it determines the surface-level syntactic analysis of text which has gone through prior morphological analysis. The process of syntactic analysis consists of three stages: morphological disambiguation, identification of clause boundaries, and identification of the syntactic functions of words. The grammatical features of words are presented in the form of tags which are attached to words. The tags indicate inflectional and derivational properties of the word and the part of speech of the word; the tags attached during the final stage of analysis indicate its syntactic functions. The underlying principle in determining both the morphological interpretation and the syntactic functions is the same: first all the possible labels are attached to the word-forms and then the ones that do not fit the context are removed by applying special rules or constraints. CG consists of hand-written rules which, by checking the context, decide whether an interpretation is correct or whether it has to be removed.

The input to the morphological disambiguator is a text that has already been morphologically analyzed. The morphological analyzer finds all possible morphological analyses for a word form; the task of the disambiguator is to decide, on the basis of the context in which the word occurs, which of the analyses suggested by the morphological analyzer is correct. During the process of disambiguation clause boundaries are also identified. Subsequently, every word in the input text is put into correspondence with a tag indicating its syntactic function. After the morphological disambiguation stage the rules for adding tags are applied to every sentence, which, on the basis of morphological information and context, attach to each word in the sentence all its possible syntactic tags. After the attachment of tags, the CG analyzer applies syntactic constraints that remove tags that do not match the context.

Thus, the work of the syntactic analyzer is entirely based on removing ambiguity. This means that already in the stage of devising rules the analyzer is able to process the sentence. If there are no rules for some homonyms, the form is left as it is, i.e. as ambiguous. In this way every sentence will always get some analysis, the difference being that some anal-

yses are more unambiguous or correct than others. Attempts are made to devise rules that will remove as few correct interpretations as possible and hence result in analyses that are as error-free as feasible.

A number of rules are clearly heuristic in nature – the rule might not be correct in 100% of cases but its proficiency rate is very high, compared to the number of errors it produces. These rules are separated into a special part of the grammar and may be excluded from the analysis process if needed.

CG seemed to be suitable for the analysis of Estonian texts because its mechanism is simple and easily implementable, it can be well adapted for the Estonian language, it is at the same time sufficiently reliable (robust) and the syntactic analysis that the Grammar outputs is suitable for various practical applications.

3 The system of adpositions in Estonian – a linguistic overview

Adpositions are invariable words that form adpositional phrases together with nominals. The adpositional phrase usually functions as an adverbial in the sentence, but can also be an attribute; the same functions can be performed by the case forms of nouns.

As mentioned earlier, adpositions in Estonian do not form a closed class. In real texts, there is, rather, a continuum comprising nouns, relational nouns and adpositions, as well as non-finite verb forms and adpositions, so that the exact boundary between word classes is difficult to define. Adpositions can combine with nominals and form three-member sets “declining” in local cases. There are three inner and three outer local cases in Estonian.

Estonian has both pre- and postpositions. Postpositions are more numerous, most of them govern genitive case on their noun complements in an adpositional phrase:

- (1) a. maja taga
house-GEN behind (location)
'behind the house'
- b. maja taha
house-GEN behind (direction)
'behind the house'
- c. maja tagant
house-GEN from behind

'from behind the house'

Some postpositions require the noun in an adpositional phrase to be in the nominative or partitive case or in the form of various oblique cases, e.g., elative or comitative:

- (2) a. päev läbi
day-NOM throughout
'all the day'
b. aastaid tagasi
year-PL-PART ago
'many years ago'
c. hommikust peale
morning-ELAT since
'since morning'

Most prepositions require the noun in the prepositional phrase to be partitive, some prepositions also govern the genitive or the oblique cases – allative, terminative, comitative or abessive:

- (3) enne õhtut
before evening-PART
'before the evening'

Many of the word-forms functioning as adpositions can function also as adverbs and construct a particle verb together with a verb. Its constituents behave much like the constituents of a particle verb in German, i.e., they are adjacent in verb-final sentences, but separated in other configurations:

- (4) a. Ma vaatan need
I-NOM look-1-SG this-PL-NOM
paberid
paper-PL-NOM
homseks ise üle.
tomorrow-SG-TRANSL self over.
'I shall look over those papers myself by tomorrow'
b. Kui sa need
If you-SG-NOM this-PL-NOM
paberid üle vaatad, siis
paper-PL-NOM over look-2-SG then
on kõik valmis
be-3-SG everything-SG-NOM ready
'If you look over those papers, then everything will be ready'

An Estonian particle verb, just like an English phrasal verb, can have its own semantics, which differs from the semantics of the simple verb, and make up its own argument structure, as in English sentences.

- (5) a. He looked over the papers in less than 10 minutes.
b. He looked over the fence and saw his neighbor.

The word-forms *look* and *over* form a phrasal verb *look over* in example (5a), but do not belong together in the same way as in example (5b), the Estonian verb *vaatama* 'to look' and adverb *üle* 'over' form a particle verb in examples (4a) and (4b), but not in example (6):

- (6) Ta vaatas üle
s/he-NOM look-3-PST over
aia ja nägi oma
fence-SG-GEN and see-3-PST own
naabrit.
neighbor-SG-PART
'S/he looked over the fence and saw his/her neighbor'

Thus, it is essential for a successful syntactic analysis of a sentence to determine correctly the part-of-speech and the syntactic function of a word-form that could be a particle (adverb) and so possibly change the meaning and the argument structure of a verb or, alternatively, be an adposition. As adpositions have to be adjacent to the nouns or noun phrases that are the constituents of the adpositional phrase, they are usually easier to detect.

4 Some problems of manual morphological (word class) annotation

During the process of manual morphological annotation at first all the possible analyses are added to each word form in the text and then a human annotator picks out the correct analysis, so the output should be 100% correct. The manual annotation of texts is essential for the development and testing of computational systems but it should also provide material for theoretical linguistic analysis. These two objectives create a discrepancy: for computational linguistic purposes every word form in the text

should have only one correct analysis, although the clear-cut boundaries between word classes are not always linguistically motivated. A linguistic analysis, on the other hand, is interested in such vague border areas and for this purpose ambiguous word forms should retain all possible analyses. As the primary goal of the manual annotation in our project was the development of computational systems, we decided to allow only one analysis per word form in a text.

During the annotation process the most trouble was caused by the idea of a predetermined set of adpositions. Annotating the real texts showed us the inadequacies of a lexicon of adpositions based on existing grammar descriptions and dictionaries and so we had to update the lexicon in parallel with the annotation process.

As mentioned before, the main problem of the adpositions is that they do not form a closed class in Estonian. Most adpositions are typical representatives of their class; they form an adpositional phrase together with a noun and are used to bear meanings close to the case-endings. Most adpositions (as well as adverbs) have emerged as products of the grammaticalization of the declensional forms of nouns or non-finite forms of verbs, a process that has gone hand in hand with the shift of meaning towards abstractness. Hence there exist word forms in texts that are undergoing this process of grammaticalization. These word forms can behave as full nouns or verbs in some contexts but in some constructions they have moved away from their initial lexical meaning and behave like relational words. This is an ongoing process in contemporary Estonian and results in word forms that are ambiguous between several word classes even in longer contexts. For example, the inessive, illative and elative case forms of the noun *käsi* 'hand' form one such source of ambiguity as they can be used both as full nouns or postpositions as exemplified in section 5.

Let us now consider the noun *pool* meaning 'half', also 'side'. Its genitive case form is *poole* and partitive *poolt*. For the following grounds we need also the allative, adessive and ablative case forms of this noun that are *poolele*, *poolele* and *poolt* respectively. There are also postpositions *pool*, *poole* and *poolt* meaning 'in/from the direction of' that require the noun in the prepositional phrase

to be in the genitive case:

- (7) a. *põhja pool*
north-GEN
'to the north of'
- b. *kodu poole*
home-GEN
'towards home'
- c. *kodu poolt*
home-GEN
'from the direction of home'

During manual morphological annotation the annotators came across usages like the following:

- (8) a. *ühel pool*
one-ADESS
'on one side'
- b. *igale poole*
every-ALLAT
'to every side, everywhere'
- c. *teiselt poolt*
other-ABLAT
'from the other side, on the other hand'

Here we have two possible morphological analyses: 1) allow the postpositions *pool*, *poole* and *poolt* to require adessive, allative or ablative case forms of the nouns in postpositional phrases and treat these constructions as postpositional phrases or 2) interpret the word forms *pool*, *poole* and *poolt* in these constructions as haplogically shortened forms of the word forms *poolele*, *poolele* and *poolt*. The latter solution is linguistically more relevant, but problematic from the point of view of automatic morphological disambiguation, because the first components of such constructions form an open set. We have chosen the first solution, mainly due to practical considerations.

In addition to the examples presented above the corpus also contained sentences where the constructions from previous examples (i.e. the word forms *pool*, *poole* or *poolt* preceded by a noun in the genitive case or the word forms *pool*, *poole* or *poolt* preceded by a noun in the allative, adessive or ablative case) as a whole behave like prepositions and require partitive case on the following noun:

- (9) a. teisel pool lauda
 other-ADESS table-PART
 'at the other side of the table'
- b. teisele poole maja
 other-ALLAT house-PART
 'to the other side of the house'
- c. hommiku poole ööd
 morning-GEN night-PART
 'in the very early morning'
- d. teiselt poolt kassautomaati
 other-ABLAT cash-desk-PART
 'from the other side of the cash desk'

Most examples of this kind contain the constructions *teisel pool*, *teisele poole* and *teiselt poolt*, but some other first components also occur in the corpus, so the construction is at least partially productive. From the point of view of theoretical linguistics, we could just as well say that we are dealing here with constructions that belong to the periphery of the adpositional phrase, but during the morphological annotation we have to give some analysis to every word form in the text. And the morphologically annotated text must be a suitable input for the syntactic analysis. How should we define the relations between these word forms? Does a preposition have an attribute? Or does one adposition bind two nouns requiring them to be in different case forms? As a practical solution the frequent constructions (e.g. *teisel pool* plus a noun in partitive, etc.) should be added to the lexicon, but more rare cases, like *hommiku poole ööd*, cause a definite error.

5 Automatic morphological disambiguation

There are 1,365 rules in the morphological disambiguation grammar (Puolakainen, 2001), 57 of which deal with adpositions. Only 1% of adpositions are unambiguous after morphological analysis and almost half of the adpositional readings are redundant.

In order to disambiguate adpositions, a general rule is employed that checks whether there are any nouns in the near context of the adposition that the adposition governs. If there are no such nouns, the word cannot be an adposition. If there is only one

such noun then the reading with a suitable case is chosen and the rest of the cases will be removed.

Adposition rules are quite successful (96.4% of adpositions become unambiguous) but certain coincidences could still cause errors. For example, in the sentence (10) the word *kohale* (adv: to one's place, to the spot; postp: over, above) is mistakenly classified as an adposition (it is really an adverb in the sentence), because in the sentence a declinable word in the genitive case precedes it:

- (10) Ma pean arsti
 I-NOM must-PRS-1SG doctor-GEN
 kohale kutsuma ...
 call-INF

'I have to call the doctor to the spot/here'

The word group *kätte, käes, käest* is also a source of difficulties in the process of analysis, and also in all other cases where decision making is based on semantic information. The postpositions *käes*, *kätte* and *käest* have developed diachronically from the illative, inessive and elative case forms of the noun *käsi* 'hand' and remain homonymous with them in contemporary Estonian. In order to decide if these word forms in a particular context belong to the class of nouns or adpositions, the syntactic and semantic relations between word forms in a particular context are taken into account and the opacity of the meaning is also considered (Kaalep et al., 2000). Thus, in example (11a) the word form *käes* is a noun and in example (11b) it is a postposition.

- (11) a. Tal olid
 he/she-ADESS have-PST-3PL
 kindad käes
 glove-NOM-PL hand-INESS/POSTP
 'He/She had gloves on.'
- b. Ta karjub valu
 he/she-NOM cry-PRS-3SG pain-GEN
 käes
 hand-INESS/POSTP
 'He/She is crying in pain.'

In order to solve the disambiguation problem for the example above and the word group *kätte*, one must create a list of phrasal verbs and use the list to decide whether a word is an adposition, adverb or noun.

The error rate of morphological disambiguation of adpositions is 4%. It is higher than the overall error rate of the Grammar, namely 2% on average.

Adpositions can also be employed in the disambiguation of nouns. Namely, the ambiguity classes nominative-genitive-partitive and genitive-partitive-short illative are very common in Estonian. If the relevant word form is part of an adpositional phrase, the government of the adposition can be used to determine the case of the word form. For example, the word form *maja* can be nominative, genitive or partitive singular; however, in the sentence (12) the postposition requires genitive case on the adjacent noun, so the word form *maja* can be successfully disambiguated.

- (12) Puu kasvab
 Tree-NOM grow-PRS-3SG
 maja kõrval
 house-NOM/GEN/PART postposition
 'The tree is growing next to the house.'

On the other hand, the application of this rule with ambiguous adpositions can cause errors. For instance, the adposition *mööda* could be either a preposition or a postposition. Therefore, in the sentence (13) the word *kooli* is mistakenly classified as a partitive noun (the word form *kooli* is ambiguous between genitive, partitive and short illative, the postposition *mööda* requires partitive case). The right decision would have been short illative.

- (13) sõitsime võimalust
 ride-PST-1PL possibility-PART
 mööda kooli
 POSTP/PREP school-GEN/PART/ILLAT
 talumeeste regedel
 farmer-PL-GEN sledge-PL-ADESS
 'If possible, we rode to school on the peasants' sledges.'

6 Syntactic disambiguation

There are 1,128 syntactic constraints in the syntactic disambiguation grammar (Müürisep, 2000), 67 of which deal with adpositional phrases. Two groups of rules are related to adpositions: rules dealing with the head of the noun phrase (or the numeral or pronoun phrase) belonging to the adpositional phrase,

and rules for determining the syntactic function of the adpositional phrase itself.

The case of a noun depends on the government of the adposition, therefore the adposition functions as the head of the adpositional phrase. Thus, the adposition is annotated with a tag indicating the syntactic function of the whole phrase, and the noun with a tag indicating whether it belongs to the pre- or postpositional phrase (<P or P>, respectively).

The head of a noun phrase in the postpositional phrase is mostly determined by government: the word with a matching case that is closest to the postposition will be analyzed as the complement of the postposition.

- (14) a. Samal ajal lähenes
 Same time approach-PST-3SG
 teist tänavat pidi veel
 other-PART street-PART along more
 üks elusolend.
 one-NOM living.being-NOM
 'Another living being was approaching along another street at the same time.'
- b. Samal-NN> ajal-ADVL lähenes-FMV
 teist-NN> tänavat-P> pidi-ADVL veel-
 ADVL üks-NN> elusolend-SUBJ

pidi (along) is a postposition that governs a noun in the partitive. In this sentence the word form *tänavat* (street-SG-PART) is labelled with P> (complement of a postposition).

This rule produced one error in a corpus of 20,000 words, namely in a sentence with a postmodifier. The postpositional complement label is removed if an unambiguous complement of a postposition has been already found and these words are not co-ordinated.

The co-ordination of two postpositional complements is also tested: if one of them is unambiguous, then the other should be analyzed as a postpositional complement too. One should be careful when treating phrases with co-ordinated premodifying attributes of a postpositional complement, e.g.,

- (15) valdade ja väikelinnade
 parish-PL-GEN and small_town-PL-GEN
 esindajate seas
 representative-PL-GEN among.

'among the representatives of parishes and small towns'

Here the word form *valdade* (parish) remained ambiguous and was labelled as the complement of a postposition and a premodifying attribute.

This phrase was the only one in the training corpus where the complement of a postposition remained ambiguous.

Also, the co-ordination rule caused one error due to an incorrectly detected clause boundary.

The rules for prepositional phrases are more difficult to compile because there are no definite restrictions on the position of the prepositional complement, and thus one cannot assume that the word closest to the preposition is the complement of the preposition. The most general rule determines the word to be the complement of the preposition if the word corresponds to the government of the preposition and in the close context there are no other potential complements in the same case. For example in the sentence (16), the preposition *läbi* 'through' governs a noun in the genitive case.

- (16) Aknatagusest
situated_beyond_window-ELAT
vaatepildist õhkus isegi läbi
view-ELAT radiate-PST-3SG even through
klaasi külma.
glass-GEN frost-PART.
'The view from the window radiated frost
even through the glass.'

If the complement of the preposition has been already found and the words are not co-ordinated, the <P tag is removed.

There are also heuristic rules that analyse words left ambiguous by previous rules – if there are two nouns in the genitive case after a preposition, then it is very unlikely that the first of them is the complement of the preposition.

Although writing rules for prepositional phrases was an awkward task, no complement of the preposition remained ambiguous in the training corpus.

Unfortunately, the determination of the syntactic function of adpositions themselves was not so successful. The adpositional phrase can function as a premodifying attribute, postmodifying attribute or adverbial.

The answer to the question whether the adpositional phrase is an attribute generally depends on semantics, but in some cases fortunately also on government. Since there is no thorough dictionary of government yet, the premodifying attribute tag is removed from adpositions with the help of heuristic rules. The rules in the main part of the grammar check whether in the right context there are nouns suitable for modification, and if there are no such nouns, the premodifying attribute tag will be removed. Heuristic rules of the third rank remove the premodifying attribute tag from the postposition if the following word is not a noun, and remove the tag from the preposition if no noun follows its complement. Heuristic rules of the fifth rank remove the premodifying attribute tag from all adpositions, except *moodi* 'like'. These rules produce 3 errors (cf. sentences 17a and 17b), but remove 76 tags correctly.

- (17) a. kes töötas Winstoni kõrval
who worked Winston-GEN next to
boksis;
box
'who worked in the box next to Winston'
b. ... luupainaja käes vaevleja ...
... nightmare-GEN in sufferer ...
'the one haunted by the nightmare'

There is also a rule in the main grammar that removed the postmodifying attribute from the adposition, if in the left context there are no nouns that could be modified. In general, an adpositional phrase functions as a postmodifying attribute more frequently than as a premodifying attribute. For this reason, a number of errors (8.5%) are produced by the heuristic rule that removes the postmodifying attribute tag from all adpositions – unfortunately, there is no other way to disambiguate adpositions. For example, one of the errors occurred in sentence (18):

- (18) ... muutus tema salajane vihkamine
... changed his secret-NOM hate-NOM
Suure Venna vastu
Big-GEN Brother-GEN against
imetluseks.
admiration-TRANSL

'his secret hate against Big Brother changed into admiration.'

In order to disambiguate such sentences, very thorough lexicons must be created containing information about noun government, and also, each individual adposition must be studied by the linguist for its suitability as an attribute. However, this research problem lies outside the scope of this paper.

The heuristic rule that removes a postmodifying attribute has a relatively high rank (6); therefore it is possible to switch the rule off when the analyzer program is used (the rules with lower ranks are applied first).

The choice between attributes and adverbials is also complicated in the case of nouns, being one of the main sources of ambiguity in the whole grammar. Table 1 presents the success-rate comparison of noun and adposition analyses.

	Recall	Precision	Unambiguity
Adpositions 1	99.43%	50.10%	67.40%
Adpositions 2	96.38%	91.95%	99.42%
Substantives, oblique cases	99.82%	39.50%	59.70%
All	98.53%	87.57%	89.54%

Table 1: Results of the parser

Here recall is defined as the ratio 'assigned appropriate labels/all appropriate labels' and precision as the ratio 'assigned appropriate labels/all assigned labels'. The percentages of unambiguous words are given in the third column. The first row presents the results for an adposition analysis without involving heuristic rules with the highest rank, while the second row describes the results for an adposition analysis with all grammar rules. The third row describes the results for substantives (excluding nominative, genitive and partitive cases) and the fourth row gives the overall results. All the results have been obtained using a corpus of 20,000 words which has been morphologically disambiguated by hand.

As can be seen from the table, the analysis of adpositions gives slightly better results than the analysis of substantives, but these results are still not as good as in the overall case.

7 Conclusions

The manual annotation of a big text corpus draws our attention to the lack of a good working definition of adpositions and to the inadequate representation of adpositions in the lexicon. Once these problems had been solved, the automatic analysis of adpositions also improved.

Nominals belonging to adpositional phrases do not cause problems at the level of syntactic analysis: the precision and recall rates of their analysis are close to 100%. At the same time the assignment of a correct syntactic analysis to the adpositional phrase as a whole can be problematic. The right choice between premodifying attribute, postmodifying attribute and adverbial analyses is as difficult as for nouns in oblique case forms. The fact that adpositional phrases seldom act as attributes in Estonian enables one to write 'robust' disambiguation rules that in most cases simply delete the attributive reading.

Further ways to improve the Grammar could involve checking adpositions word by word in the text corpus to specify the cases in which adpositional phrase can function as an attribute and writing specific new rules for these cases.

References

- Mati Ereht, editor. 2003. *Estonian Language*, volume 1 of *Linguistica Uralica Supplementary Series*. Estonian Academy Publishers, Tallinn.
- Tiit Hennoste, Mare Koit, Tiit Roosmaa, and Madis Saluveer. 1998. Structure and usage of the tartu university corpus of written estonian. *International Journal of Corpus Linguistics*, 3(2):279–304.
- Heiki-Jaan Kaalep, Kadri Muischnek, Kaili Müürisep, Andriela Rääbis, and Külli Habicht. 2000. Kas tegelik tekst allub eesti keele morfoloogilistele kirjeldustele? Eesti keele testkorpuse morfosüntaktilise märgendamise kogemusest. *Keel ja Kirjandus*, (9):623–633.
- Heiki-Jaan Kaalep. 1996. ESTMORF. A Morphological Analyzer for Estonian. In Haldur Õim, editor, *Estonian in the Changing World*, pages 43–98. University of Tartu, Tartu.
- Fred Karlsson, Atro Voutilainen, Juha Heikkilä, and Arto Anttila. 1995. *Constraint Grammar: a Language Independent System for Parsing Unrestricted Text*. Mouton de Gruyter, Berlin and New York.

Kaili Müürisep. 2000. *Eesti keele arvutigrammatika: süntaks*. Dissertationes Mathematicae Universitatis Tartuensis 22. Tartu ülikooli kirjastus, Tartu.

Tiina Puolakainen. 2001. *Eesti keele arvutigrammatika: morfoloogiline ühestamine*. Dissertationes Mathematicae Universitatis Tartuensis 27. Tartu ülikooli kirjastus, Tartu.