

The case of the Estonian argument structure constructions containing infinitival arguments

Heete Sahkai¹, Kertu Saul^{1,2}, Raili Pool^{1,2}, Jelena Kallas¹, Katrin Tsepelina^{1,2}, Ene Vainik¹, Geda Paulsen¹, Arvi Tavast¹, Kristina Koppel¹
Institute of the Estonian Language¹, University of Tartu²

21st Annual Conference of Applied Linguistics 2024

April 19, 2024

Outline

Goal: Connecting lexeme-specific government patterns with argument structure constructions (ASCs) in the EKI Combined Dictionary

- Preliminary considerations
- The case of Estonian infinitival arguments: a difficulty for learners
- Constructicographic representation of ASCs
- Data sources
- Methods for corpus-based identification of constructions and for the extraction of instances of constructions
- Methods for corpus-based identification of proficiency levels

Preliminary considerations

Connecting lexeme-specific government patterns with ASCs:

- Economy
- More systematic coverage
- Pedagogical benefits:
 - guiding learners through semantically and syntactically similar words
 - helping teachers to compile teaching materials that support the acquisition of meaningful patterns
- Representing the productivity of ASCs

A case study: *ma-* and *da-* infinitival arguments

ujuma / *ujuda* ‘to swim’

Ta õpib uju-ma.

3SG learn.3SG swim-MA

‘S/he is learning to swim.’

Ta armastab uju-da.

3SG love.3SG swim-DA

‘S/he loves to swim.’

Infinitival arguments in Estonian L2 learner data: non-target-like usage examples

- The use of *ma*-infinitive with verbs requiring the *da*-infinitive
Veel ma **tahaks ostma** uut köögimööblit. 'I would like to buy new kitchen furniture.' (Adult learner, B1 level)
- The use of *da*-infinitive with verbs requiring the *ma*-infinitive
Homme **pean minna** arsti juurde. 'Tomorrow I have to go to the doctor.' (Adult learner, B1 level)
- The use of the *ma*- or *da*-infinitive with verbs that require an (event) noun as the complement of the main verb
Kui sulle meeldib inimene ja sa tahad **jätkata suhtlema**... 'If you like a person and you want to continue communicating...' (Young learner, B2 level)

Current state in the Combined Dictionary: infinitival government

Selected datasets: 1

keyword

☐ government is mida teha

☐ part of speech is tegusõna, verb

Results found:56

1 aitama (et)
eki, ety, gal, les

1 armastama (et)
eki, gal, les

1 eelistama (et)
eki, mat

1 häbenema (et)
eki, les

1 igatsema (et)
eki, ety, gal, les

1 ihkama (et)
eki, les

1 jaksama (et)
eki, les

1 jõudma (et)
eki, gal, les

1 julgema (et)
eki, les

1 kartma (et)
eki, les

1 katsuma (et)
eki, gal, les

1 kavas olema (et)
eki

armastama Join 06.03.2024

eesti Morph code Vocal form armasta[ma Gender Word type Aspect

IDENTIFIERS	W158021 L1148873 M14045	
LEXEME TAG	+	prantsuse tehtud , süno valmis
COMPLEXITY	simple/detailed	
LEXEME WEIGHT	1	
PART OF SPEECH	+	v - tegusõna, verb
SEMANTIC TYPE	+	VERB_tegevus - tegevus/protsess üldiselt, mingit tegevust sooritama (nt lööma)
DEFINITION	+	et • midagi meeleldi (ning seetõttu sageli) tegema [eki] • kui sa midagi armastad , siis see asi või tegevus meeldib sulle väga [eki]
MEANING ATTRIBUTES	CONCEPT_ID 119569 CONCEPT_ID 1224	
GOVERNMENT	+	mida* mida teha
USAGE EXAMPLE	+	Salme on kogu elu armastanud kududa Kihnu tekke. Memm armastas rääkida igasuguseid lugusid. Ma armastan muusikat. Ta armastab laulda.
COLLOCATIONS	V	V_Vda FREQUENCY: 3558.0 SCORE: 29.6067
	armastab öelda	ütlemata [0.8] 292.0 6.9629 Kasutame oma halle ajurakke, nagu kuulus detektiiv armastas öelda.
	armastab rääkida	rääkima [0.8] 191.0 7.9506
	armastab rõhutada	rõhutama [0.8] 103.0 8.9452

Goal:

- Adding entries for ASCs
- Connecting them to lexeme entries
- Showing full clausal patterns in lexeme entries

Distribution of the *ma*- and *da*-infinitival arguments

Data and analysis

- Verbs with *ma*- and *da*-infinitival arguments in the Combined Dictionary: semantic groupings and clausal patterns
- Using queries based on the clausal patterns to extract additional data from the Balanced Corpus and to validate the groupings
- Penjam (2008)

The complementarity of the *ma*- and *da*-infinitives: prototypical meanings

V_{motion/change} + **V-MA**_{goal/purpose}

V_{state} + **V-DA**_{situation}

Both suffixes originate in lative case forms of deverbal nominalising suffixes (Laanest 1975) but this origin is more transparent in *ma*-infinitives

Semantic oppositions

Change (MA) vs. state (DA)

Õpin lugema 'I'm learning to read' vs. *Tahan lugeda* 'I want to read'

More causative (MA) vs. less causative (DA)

Sundisin ta lugema 'I forced her to read' vs. *Aitasin tal lugeda* 'I helped her to read'

Inchoative (MA) vs. continuous/habitual (DA)

Hakkan lugema 'I start to read' vs. *Tavatsen lugeda* 'I'm used to reading'

Deontic necessity (MA) vs. dynamic modality and deontic possibility (DA)

Pean lugema 'I must read' vs. *Saan/võin/tohin lugeda* 'I can read'

Active (MA) vs. active/passive (DA)

Kleit sobib kanda 'The dress is suitable to be worn'

The complementarity of the *ma-* and *da-*infinitives: syntactic distribution

- ma-* and *da-*infinitival arguments occur in different syntactic patterns

Type of syntactic pattern	MA	DA
Infinitival complement, subject control <i>Ma õpin lugema</i> 'I'm learning to read', <i>Ma tahan lugeda</i> 'I want to read'	YES	YES
Infinitival complement, object control <i>Ma õpetan teda lugema</i> 'I'm teaching him/her to read'	YES	NO
Infinitival complement, oblique control <i>Ma aital tal lugeda</i> 'I'm helping him/her to read'	NO	YES
Infinitival subject, no/oblique control <i>(Sul) tasub lugeda</i> 'It's worth while (for you) to read'	NO	YES
Infinitival copular subject, no/oblique control <i>Lugeda on huvitav</i> , <i>Mul on huvitav lugeda</i> 'It's interesting (for me) to read'	NO	YES

- Each syntactic pattern occurs with (sets of) semantically related main verbs and can be analysed as a (polysemous) ASC or a set of ASCs

An example

Da-infinitival complements controlled by the main verb subject occur with semantically related sets of main verbs:

Verbs denoting a **mental disposition** regarding a possible action (Van Valin & La Polla 1997:479): *tahtma* 'want', *kavatsema* 'intend', *otsustama* 'decide', *taipama* 'realise', *unustama* 'forget'

Affective disposition verbs: *kartma* 'be afraid', *häbenema* 'be ashamed', *armastama* 'love'

Ability verbs: *suutma* 'be able', *oskama* 'can'

Conative verbs: *püüdma*, *proovima* 'try'

Speech act verbs: *lubama* 'promise', *ähvardama* 'threaten'

- **Mental disposition construction**
- **Affective disposition construction**
- **Ability construction**
- **Conative construction**

Constructicographic representation of the Mental Disposition Construction

Constructicon entry elements

- type (schematic, partially schematic, lexical cxn)
- name
- form
- constituent and dependency structure
 - morphological form of the components
 - syntactic function
 - phrasal type
- meaning
- productivity description
- language proficiency level
- frequency information
- regular expression for corpus query

Mental Disposition Construction entry elements (1)

Type:	Argument Structure Construction	
Name:	Mental Disposition Construction	
Form:	who VERB to do what (user friendly)	
	[NPnom _i]	[V _i [VP-DA]]
	Subject	Complement
Role:	Actor	Action
Meaning:	The actor has a mental disposition regarding the action	

Mental Disposition Construction entry elements (2)

Verbs filling the slot of the main verb with links to the corresponding verb entries: *tahtma* 'want' A1, *igatsema* 'long' B2...

Proficiency level of the construction: A1 (the lowest level where a construction is used with any verb)

Proficiency level of verb-construction combination (the lowest level where a particular verb is used in this construction)

Frequency of the construction

Frequency or association strength of particular verb-construction combinations

Examples: (authentic) examples picked by the lexicographer

CQL for corpus query

Data sources for compiling Argument Structure Construction entries

- EKI Combined Dictionary and government dictionaries
- PropBank for Estonian (under development)
- Existing studies, e.g., Rätsep (1978), Penjam (2008)
- Corpora

Methods for corpus-based identification and extraction of ASCs

- Estonian National Corpora
multi-layer: EstNLTK's StanzaSyntaxTagger
- **dependency relations + morphological annotation**

Identification:

- Word-based
- Construction-based
- Explorative, e.g. n-grams

Methods for corpus-based identification of proficiency levels

- Estonian L2 learner corpora
- Estonian L2 textbook corpora

Tools: Sketch Engine, [EMMA](#), [ELLE](#)

- identification of constructions (A1-C1) in level-specific (sub)corpora
- manual and computational analysis of the data

Eesti Keele Instituut

ÕPETAJA TÖÖRIISTAD

SõnavaraGrammatikaKasutusolukorradÕppeteksti hindamineÕppevaraProjektist

Grammatika

Grammatika osa sisaldab noore ja täiskasvanud keeleõppija grammatikapädevuse kirjeldust. Noore keeleõppija grammatikapädevuse kirjeldus on koostatud tasemetele eelA1–B2. Tasemekirjelduste vastavused keelepädevusega on järgmised: eelA1 = koolieelikud (6–7-aastased), A1 = I kooliaste (1–3. klass), A2 = II kooliaste (4.–6. klass), B1 = III kooliaste (7.–9. klass), B2 = gümnaasium (10.–12. klass). Täiskasvanud keeleõppija grammatikapädevuse kirjeldus on koostatud tasemetele A1–C1.

Grammatikapädevuse kirjeldus on üles ehitatud laienevalt: lihtsamast keerulisemani. Sama kategooria on sageli kirjeldatud mitmel keeleoskustasemel (*oskab-lauses*), kuid kasutusfunktsioone ja piiranguid (kui tihti õppija eksib ja milles eksimus seisneb) on täpsustatud vastavalt keeleoskuse arengule. C1-tasemel kirjeldatakse ainult need vormi- ja sõnamoodustuse teemad, mis madalamatel tasemetel veel ei esildunud.

Keeleõppija:

☐ noor keeleõppija

☒ täiskasvanu

Keeletase:

☐ A1

☐ A2

☐ B1

☐ B2

☐ C1

Peakategooria

FRAASIMOODUSTUS

Alamkategooriad

TEGUSÕNAFRAAS

laiend

tegevusnimi

+ Lisa peakategooria

Otsingu tulemused (13)

Peakategooria	Alamkategooria	Tase	Kirjeldus
FRAASIMOODUSTUS	TEGUSÕNAFRAAS	A1	tegusõna koos <i>da</i>-tegevusnimega Oskab kasutada mõne sagedase tasemekohase tegusõna laiendina <i>da</i>-tegevusnime (nt <i>tahtma</i>, <i>oskama</i>, <i>saama</i> + <i>-da</i>). Tegusõna vormide moodustamisel ja tegevusnime valikul eksib sageli. Näited: <i>Ma tahan magada. Mulle meeldib lugeda. Ma oskan ujuda.</i> Tunnused: laiend <i>tegevusnimi</i>
FRAASIMOODUSTUS	TEGUSÕNAFRAAS	A1	tegusõna koos <i>ma</i>-tegevusnimega Oskab mõne sagedase tasemekohase tegusõna laiendina kasutada <i>ma</i>-tegevusnime (nt <i>minema</i>, <i>pidama</i>). Tegusõna vormide moodustamisel ja tegevusnime valikul eksib sageli. Näited: <i>Ma lähen magama. Tule sööma!</i> Tunnused: laiend <i>tegevusnimi</i>

[Estonian L2 Grammar Profile Description of Infinitival constructions](#)

Open questions

- Distinguishing auxiliary constructions, ASCs and adjunct constructions
- Distinguishing constructional variation, polysemy and subconstructions
- User-friendly representation of the form of constructions
- Identifying and representing constructional meaning
- Identifying and representing constructional productivity
- ...



References

Laanest, A. 1972. Sissejuhatus läänemeresoome keeltesse. Tallinn: Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut

Penjam, Pille 2008. Eesti kirjakeele *da-* ja *ma-*infinitiiviga konstruktsioonid. Tartu: Tartu Ülikooli Kirjastus.

Rätsep, H. 1978. Eesti keele lihtlausetes tüübid. Tallinn: Valgus.

Van Valin, R. D., Jr. & R. D. LaPolla 1997. Syntax. Structure, meaning and function. Cambridge: Cambridge University Press