EURALEX XIX

Congress of the European Association for Lexicography

Lexicography for inclusion

7-11 September 2021

Ramada Plaza Thraki
Alexandroupolis, Greece

www.euralex2020.gr

Proceedings Book
Volume 1

Edited by Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

EURALEX Proceedings

ISSN 2521-7100

ISBN 978-618-85138-1-5

Edited by: Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

English Language Proofreading: Lydia Mitits and Spyridon Kiosses

Technical Editor: Kyriakos Zagliveris

2020 Edition

A Typology of Lexical Ambiforms in Estonian

Vainik E., Paulsen G., Lohk A.

Institute of the Estonian Language

Abstract

The present study aims to elaborate an overall outline of the areas that give rise to PoS ambiguity in Estonian. The analysis is based on a database consisting of ca 3500 ambiguous units. Our goal is to map the problematic areas, analyse the processes behind the lexical versatility, and provide a typology of ambiguous forms (the ambiforms) for the lexicographic use. The proposed typology is based on bi- and unidirectional PoS combinations. As a result of the analysis, we show how the lexical confluence relations exhibit a network-like interaction of the traditional PoS categories. The typology of ambiforms is expected to have both theoretical and practical implications – from the perspective of the former, the topic of lexical ambiguity will be set in the modern linguistic and lexicographic frame, and from the angle of applicability, the results will support the lexicographers as the creators of lexicographic (root) databases and the developers of language technology systems analysing corpus data.

Keywords: parts of speech; lexical ambiguity; Estonian language

1 Introduction

The contemporary lexicographic processes are characterised by the tendency to data unification, which sets new demands on and urges rethinking of the part of speech (PoS) categorisation issues. Within Estonian lexicography, the procedure of part-of-speech tagging is considered unavoidable both in the compilation of new dictionaries and in the aggregation of the existing ones into lexical databases. The PoS categorisation is a part of the data model in Ekilex –the newest dictionary writing system and lexicographic root-database of The Institute of the Estonian Language. Explicit marking of word classes of every lexical entry is the ultimate goal in the output of the lexicographic resources gathered into Ekilex – the Combined Dictionary of Estonian (DicEst) 2020 and the Language portal Sõnaveeb[1] ("Wordweb 2019"; see Tavast et al. 2018; Koppel et al. 2019).

Generally, the word class categorisation is not a complicated task for the lexicographers, but in ambiguous cases, the decision making can be rather difficult (see Paulsen et al. 2019). In Estonian, as in many other languages, the boundaries between word classes are not always clear. One of the forces behind the word class ambiguity is conversional transposition, expressed in particular in the adjective-noun direction (see Vare 2006). As a morphologically rich language, Estonian also has a number of inflected forms that tend to move their basic lexical categorial status to another (see e.g. Grünthal 2003). The inflected word forms emerge as candidates for autonomous dictionary entries, often in a different PoS category than the base word. There are, continuously, plenty of word forms in a transition stage (Erelt et al. 2017).

For these kinds of words, i.e. words or word forms that may be interpreted as belonging to more than one word class, we use the general term *ambiform*. The ambiforms, if not handled properly, cause problems while integrating different lexicographic resources into larger databases and provide inaccurate results in the corpus processing systems. Therefore, there is a need to discover the types of potential ambiforms and to create, if possible, standardised procedures to handle them in lexicographic databases as well as for disambiguation of corpus data.

The aim of the present study is to elaborate an overall outline of the areas that give rise to PoS ambiguity in Estonian, based on the existing lexical databases and the data collected during lexicographic work and in the metalexicographic study carried out among Estonian lexicographers (Paulsen et al. 2019; Paulsen et al. 2020). The quantitative and qualitative analysis is based on a database consisting of ca 3500 ambiguous units. In this study, we map the problematic areas, analyse the processes behind the linguistic changes, and provide a typology of ambiguous forms for the lexicographic use. Validation of this typology on corpus data will be the next step in our research.

We expect the results of this study to be useful not only for Estonian lexicographers but also for other languages with rich morphology, especially from the point of view of the objective to move towards integrated lexicographic resources and harmonized standards in the lexicographic description in Europe (Pedersen et al. 2018). Another field of applicability of the typology of ambiforms are the automatic word sense disambiguation and morphological segmentation systems. Today, there is a strive for exploitation of the data of rich lexical databases for the needs of diverse language technological applications.

The article is organized as follows: the theoretical background and a basic outline of the prototypical Estonian PoS categories are presented in Section 2. Section 3 gives an overview of the content and organisation of the database; it also explains the methodological solutions in the analysis of the ambiforms. The typology of ambiforms, based on bi- and unidirectional PoS combinations, is proposed in Section 4. In section 5, the results are summarised and discussed.

[1] https://sonaveeb.ee/

2 Background

2.1 Lexical Categories and Applied Linguistics

Substantially, there are two types of approaches to linguistic categories: the classes of natural language can, by nature, be seen as discrete (classical categories) or prototype-based, graded ones.[2] The definitions of word class categories typically combine semantic, morphological, syntactic, and sometimes even pragmatic information; the categorisation being in essence language-specific. There are hence no universal parameters to specify the boundaries of PoS and the actual criteria depend largely on the properties of the language under consideration and the aims of the classification. Word classes are often also divided into closed and open classes, depending on the ability to admit new members and to convey independent meanings. The organisation of lexemes is affected by homonymy, polysemy,[3] and derivational relations; due to the dynamic processes shaping language, words or word forms move from one class to another. The category change in language is not always easily distinguishable, however, the reuse of a linguistic unit in different functions can be seen as an intrinsic feature of language, favouring many-to-many relationships.

There are many linguists that have questioned the sufficiency of lexical categories to capture the grammatical behaviour of words (e.g. Culicover 1999; Croft 2001; Taylor 2012; Smith 2015). However, PoS as a categorial frame is not significant only from the theoretical point of view, these concepts are fundamental in applied linguistics as lexicography and language technology; the theoretical problems related to lexical categories are even more exigent in part-of-speech tagging and word sense disambiguation procedures. There is, for example, a set of universal tags for PoS (UPOS) developed in order to enable cross-linguistically consistent treebank annotation (see e.g de Marnfette et al. 2014). Expanding flexibly the coverage of the limited number of UPOS labels and using an additional XPOS tag for the language specific categories seems to be a strategy to handle the less prototypic or ambiguous cases.[4] Such a strategy works well in the case of corpus tokens, which are surrounded by an immediate context and are, in principle, disambiguable. Several methods – either rule-based, probabilistic or neural – have been invented for morphological disambiguation (see e.g. Quecedo 2019 for further references). However, attributing a POS label to a decontextualized dictionary entry is a different task, which, inevitably, lies on a generalisation made over the correct analyses of tokens in the corpora.

When creating corpora and lexical databases over Estonian, the crucial information for distinguishing between different realisations of ambiforms is word class affiliation; incorrect or absent PoS tagging yields incorrect automatic analysis (see Koppel 2020: 62). The fluidity of the PoS-boundaries in Estonian has been explored to some extent from the lexicographic point of view. Based on the experience with compilation of the Explanatory Dictionary of Estonian (EKSS), Karelson (2005) estimates the sets of ambiguous cases to cluster around the (dominating) classes – noun, adjective, adverb, interjection, numeral, verb. Habicht et al. (2011) discuss the challenge associated with PoS subdivision by the example of adverbs (verbal particles, modal adverbs, proadverbs) in the analysis of the PoS annotation problems of Old Written Estonian lexis. A metalexicographic survey mapping lexicographers' experiences concerning PoS categorisation (Paulsen et al. 2019: 327–329; Paulsen et al. 2020) points to the following three most difficult pairs of PoS: noun-adjective, noun-adverb and noun-adposition. Hence, there are numerous ambiguous forms in Estonian with the possibility of more than one interpretation.

2.2 Estonian Parts of Speech in a Nutshell

Estonian is a language with rich morphology; it has both nominal and verbal inflection and derivation, and in addition productive compounding. Estonian words can be divided into four main classes by their morphological behaviour: (1) words that inflect for mood, time and person (verbs), (2) words that inflect for case and number including for grammatical cases, i.e. nominative, genitive and partitive (nominals, i.e. nouns, adjectives, numerals, and pronouns), (3) words that inflect in (some) semantic cases, but have no grammatical case forms (some adverb types and some adpositions), (4) words that have no inflectional forms (some adverb types and adpositions, conjunctions, interjections) (Viitso 2003: 32). From the point of view of PoS border areas, the exceptions of this division are of most interest – for instance, most adjectives have in addition to nominal inflection also forms for degrees of comparison, some atypical adjectives do not inflect, and some adverbs and adpositions may come in (series of) semantic case forms.

The morphological form of a word conveys information about its syntactic and semantic characteristics. In Estonian, the nominals decline in 14 cases, in singular and plural: three grammatical (nominative, genitive, partitive) and eleven semantic (illative, inessive, elative, allative, adessive, ablative, translative, terminative, essive, abessive, comitative) cases. In the group of nominals, nouns and adjectives form noun phrases that function as arguments of the predicate (subject, object, predicative). Nouns are the heads of noun phrases, and adjectives modify the noun, agreeing with its head in case and number (*suur-te-st tera-de-st* [big-PL-ELA grain-PL-ELA] "from big grains"). However, there are some systematic exceptions to this rule: adjective in the attributive position agrees only in number in four cases, terminative, essive, abessive and comitative, being marked with the genitive case (*suur-te tera-de-ga* [big-PL.GEN grain-PL-COM]

[2] An example of the classical categorization would be e.g. Chomsky's (1974) feature-based approach to lexical categories using a set of internal features (+/−N, +/−V); the category "verb" can be explained by the absence of the property "noun": [−N, +V]. The prototype-based categorization can be illustrated by the famous example of Rosch (1978): there are differences in how exactly different kinds of birds correspond to the concept of "bird" (the sparrow is in certain respect "birdier" than the penguin).

[3] Polysems are the elements with the same form and etymology but different meanings, the units with different etymology are homonyms.

[4] https://universaldependencies.org/u/pos/ [23.07.2020]

"with big grains").
The specific property of numerals (e.g. *kaks* "two"*, teine* "second", *neljandik* "a quarter") is that they refer to the numeral quantity and are typically used as the head of quantifying phrase with nominal complement in partitive case (*kaks last* [two kid-PART] "two kids"). Pronouns in Estonian share the syntactic properties of nouns, adjectives, and numerals, with emptier semantics.
Adverbs modify verb phrases, adjectives, or whole sentences. The adpositional phrases (both pre- and postpositions) are often used (parallelly) with nominal cases. Adpositions are a syntactically dependent word class, they are grammatical heads and determine the position and case form of the nominals participating in adpositional phrases. Conjunctions in turn play no role in the main clause structure and serve a bridging function of construals, connecting words, phrases, or clauses.
The interjection is an exceptional PoS in the Estonian lexical system. Due to their independence of the main clause structure, interjections are sometimes treated as a type of a sentence rather than a word (Erelt 2017: 61). A special case of interjections are the expressive or ideophonic (involving both onomatopoetic and descriptive) words – the irregular/abnormal category as opposed to the "neutral vocabulary", constituting a noticeable part of the Estonian vocabulary that is difficult to describe and categorise (Mikone 2002; 2001: 223).
The verb in Estonian has finite forms (occurring as predicates or auxiliary components of complex predicates) and non-finite forms. The non-finite forms occur in complex predicates with a finite form (past participles); in less verb-like functions, the non-finite forms appear also as subjects and objects (infinitive), as attributes and predicatives (participles), and as adverbials (supines and gerund). There is one infinitive (*luge-da* "to read") and one gerund – the inessive case form of the infinitive (*luge-des* "while reading"). Participles inflect for voice and tense, present participles also for case and number (*luge-va-te-ga* [read-PTC-PL-COM] "with the ones that read". Supines are inflected for voice and case, the personal supine is inflected for five cases but not for number (e.g. *luge-ma-st* [read-SUP-ELA] "from reading"). Estonian has certain verbal nouns close to non-finite forms: agent nouns (*luge-ja* "reader"; *luge-nu* "one who read"), patient nouns (*loe-tu* "something that was read"), and action nouns (*luge-mine* "reading") (Viitso 2003: 52).

3 The Database of Ambiforms

The analysis of PoS border areas grounds on the database of ca 3500 ambiforms. The different sources of data are presented in Table 1. Most of the data derive from the morphological database of Estonian (MAB).[5] Another systematic source of the ambiguous forms is the database of The Dictionary of Estonian (ES2019), where the items specified as subheadwords[6] (and, thus, missing a PoS label) were retrieved.[7] One of the largest sources is also the file of notes taken by Geda Paulsen in the course of compiling the Estonian collocation dictionary (ECD). No duplicates of ambiforms were allowed.

Source	N	%
Morphological database	2385,00	68,07%
Subheadwords from the dictionary ES2019	494,00	14,10%
Exceprts from the collocation dictionary (ECD)	447,00	12,76%
Metalexicographic study	124,00	3,54%
Literature	42,00	1,20%
Other	12,00	0,34%
Total:	**3504**	**100%**

Table 1: Constitution of the database.

The database (MS Access) comprises related tables of linguistic information. One of them records the ambiforms, i.e. the linguistic expressions, which's categorization is not straightforward and may cause PoS disambiguation problems (e.g. *asjata* "needless(ly)"). The central table records the different interpretations of ambiforms in terms of PoS (e.g. *asjata* – adjective; *asjata* – adverb). Yet the contexts giving rise to those different interpretations are stored in a separate but related table.
Further descriptors can be added to every table (containing the ambiforms, the interpretations, the contexts, or labels) targeting the aspects that are specific to or relevant for that particular object of description. For example, we suggest that the very specific PoS labels should be categorised into more general groups that would represent the basic level PoS categories (e.g. both prepositions and postpositions can be subsumed under the label adpositions). Another example of further descriptors would be a typology of contexts. It is possible to create a set of tags (pointing e.g. to the syntactic parameters) that would characterise the contexts in general terms and reveal, thus, their commonalities. Yet another further classification could be the typology of ambiforms themselves. The purpose of the present study is to propose such a typology.

[5] The database unifies the morphological info of different dictionaries compiled by different authors at the Institute of the Estonian Language. At least part of the multiple markings is due to the fact that different sources have alternative markings. Importantly, the material of MAB is mostly based on paper dictionaries that have excluded a large amount of word (forms).
[6] In the ES2019, the inflected forms detaching from their base forms (situating on an intermediate stage in their respective grammaticalization-lexicalisation processes) are tagged as subheadwords instead of separate independent headwords; the PoS tag of the subheadword is, in this case, unmarked (see Langemets et al. 2018: 948–950).
[7] We thank Ülle Viks for both excerpts.

4 Results

In broad terms, the material demonstrates ambiguity in two respects: i) interpretability of some forms as belonging to different parts of speech, and ii) ambiguity in respect of whether and in which conditions a lexical unit should be treated as a proper headword in the dictionary in its own rights (see e.g. Karelson 2005; Blensenius & Martens 2019). The typology reported here generalises information about the interpretability of the ambiforms in terms of their PoS categorisation, only. The aspect of the entrenchment of the forms as potential new headwords of a dictionary will be tackled in another study focusing on the distributions of the ambiforms as compared to the behaviour of the ordinary, non-entrenched distribution of case forms.

The essence of present typology comprises combinations of PoS categorisations that can occur as the interpretations of an ambiform. The data about combinations in or analysis originates in two sources. First, the data imported from the MAB (N=2385) was provided with the labels of combining PoS categories. These non-coincidental markings were further inherited from the numerous aggregated dictionaries. Second, our total list of 3504 ambiforms was subjected to automatic morphological analysis[8]and the interpretations including different PoS labels were marked as potential PoS combinations. We decided to adopt the same PoS categories and labels as in the morphological database and used by the automatic morphological (morph-) analysis.[9]

Altogether 33 ambiform combinations by two, 21 by three, and 5 by four tags occurred, the majority (94%) of these are biforms by nature (i.e. the forms with two interpretations in terms of PoS categorisation). Table 2 presents the most prominent combinations of PoS by two, which are, basically, the types of ambiforms that will be described more closely in the typology below.

Directionality	No	Notification	Donor	Target	CMC	Morph-analyser (Total = 3504)	Morphological database (Total = 2385)
Bidirectional	1.1	A<>S	adjective/noun	noun/adjective		40,20%	49,10%
	1.2	D<>A	adverb/adjective	adjective/adverb	yes	8,34%	14,21%
	1.3	D<>S	adverb/noun	noun/adverb	yes	8,06%	3,86%
	1.4	I<>D	interjection/adverb	adverb/interjection		1,51%	6,54%
	1.5	N<>P	numeral/pronoun	pronoun/numeral		0,29%	0,50%
Unidirectional	2.1	V >S	verb	noun		12%	0,13%
	2.2	V >A	verb	adjective		7,77%	0,21%
	2.3	S >K	noun	adposition	yes	3,57%	0,34%
	2.4	K >D	adposition	adverb		3,11%	6,29%
	2.5	V >D	verb	adverb	yes	2,71%	0,13%
	2.6	I >S	interjection	noun	yes	2,57%	3,69%
	2.7	P >S	pronoun	noun		1,06%	0,63%
	2.8	N >S	numeral	noun		0,80%	1,30%
	2.9	P >A	pronoun	adjective		0,37%	0,63%
	2.10	J >D	conjunction	adverb		0,29%	0,50%
	[…]	**Others**				**7,57%**	**12,02%**
						100%	**100%**

Note: CMC – change of morphological class

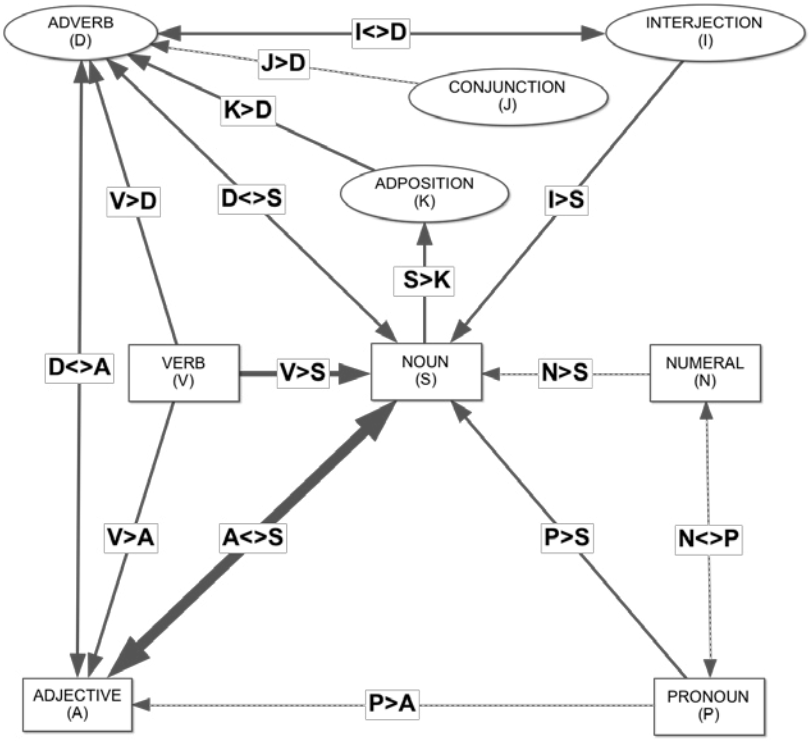
Table 2: Combinations of PoS and directions of biforms.

One can observe that some PoS categories occur in multiple biforms and are more prone to alter their interpretation than others. For example, nouns participate in 8 combinations, adverbs in 6, adjectives in 4, etc. Figure 1 presents the set of combinations as a network of PoS categories where the biforms occur as the connections. We explain the occurrence of some biforms as a shift in PoS categorisation. The shift can be seen as having a direction from a "donor" (i.e. original or dominant) PoS category to a "target" PoS category (i.e. the alternative interpretation). The incoming arrows can be described as the processes of adjectivisation, adverbisation, nominalisation, etc, respectively. Figure 1 demonstrates that some categories occur purely in the role of donors (interjections, conjunctions, verbs) while most of them can take both the roles of a donor and a target. According to the donor-target directionality, the types of biforms were further classified into bidirectional and unidirectional ones. This classification, as well as the roles of donor and target, are also presented in Table 2 together with the marking ("yes") of cases where a shift across the boundary of the main morphological class (of inflected vs non-inflected words) occurs.

The descriptions of ambiform types below involve brief observations about the morphological and syntactic aspects explaining the mechanisms of the emergence of biforms. In addition, the cases of systematic polysemy and the productivity of the patterns are commented, and some suggestions are given about handling the biforms in a dictionary. For lack of space, we do not comment nor present examples of the cases where homonymy causes the multiple interpretations.

[8] https://estnltk.github.io/estnltk/1.2/tutorials/morf_tables.html#postag-table [19.05.2020].

[9] A – Adjective; D – Adverb; G – Genitive attribute (indeclinable adjective); H – proper noun; I – Interjection; J – Conjunction; K – Adposition; N – Cardinal numeral; P – Pronoun; S – Noun; V – Verb (see the link in the previous footnote).



Legend: Rectangles – inflected; Ovals – uninflected; Bold – proportion of cases[10]; Arrow – direction.

Figure 1: Network of PoS categories as connected by biforms.

5 The Typology of Biforms

The numeration of the biforms in the typology below follows a) directionality; b) prominence (see Table 2); the subpartition in the following analysis of biform types adheres to the type numbers as presented in Table 2.

5.1 Bidirectional Types

1.1 A<>S [adjective<>noun]

The A<>S biform displays productive patterns of cross-using the two PoS categories: every adjective can employ the same syntactic functions as nouns (i.e. occur as a subject, object, or predicative), due to ellipsis, and some nouns can be used as a modifier. This type represents a two-way relation: the adjectives undergo nominalisation and the nouns undergo adjectivisation. Some of A<>S biforms are entrenched to the extent of inseparability on the scale pan between adjective and noun and can occasionally be tagged with both classes in DicEst.[11] As such, the A>S biform represents a case of systematic polysemy (Langemets 2010) (typically, QUALITY – CARRIER OF THE QUALITY), based on metonymy, and the sense menu of the dictionary reflects both senses. Also, the S>A direction reflects a metonymic relationship, where a property stands for its carrier.

(1)	A>S	*kallim*	comparative form of	“dear; expensive” “the loved one”
		rase	“pregnant”	“a gravid woman”
		loll	“stupid”	“fool”
		sinine	“blue”	“blue colour”
		vaimulik	“cleric”	“clergyman”
	S>A	*lemmik*	“favourite thing”	“favourite, dearest”
		koer	“dog”	“naughty, frisky”
		pull	“bull”	“cool, terrific”
		räbal	“rug”	“cheesy”

[10] The proportion is given according to the results of morph-analyser, see Table 2.

[11] In the analyses below, we compare our ambiform types to PoS tags in the most recent and comprehensive dictionary (regarding also the PoS marking), DicEst.

1.2 D<>A [adverb<>adjective]

This biform subsumes a subclass of atypical indeclinable adjectives that tends to occur in contexts typical for both adjectives and adverbs. It is possible to use the same D<>A biform as a universal modifier – for instance, a such biform in the position of predicative is indistinguishable from a modifier of state/position in a sentence like *Ta õlad on längus* "His shoulders are dropped". The D<>A biforms are mostly used in the informal register; they convey expressive semantics and their forms are often ideophonic, i.e. phonologically motivated (see Mikone 2001, Kasik 2015: 77). There are specific suffixes deriving D<>A biforms (*-s -kil, -li, -il, -vel, -vil*, see examples 2–7) and a sufixoid (*-võitu*). Another kind of ideophonic word formation is (partial) reduplication (see 8–9).

(2) *kiivas* "catawampus, aslant"
(3) *krussis* "curly"
(4) *laokil* "uncared for"
(5) *purjus* "drunken"
(6) *lömmis* "crumpled"
(7) *löntis* "saggy"
(8) *tippentoppen* "tiptop"
(9) *triksistraksis* "ready to go"

Some D<>A examples could be analysed also as (locative) case forms of existent nouns: *ähmi-s* "excited" [flap-INE], *küüru-s* "crooked" [hump-INE]. In some cases, a base noun (e.g. *küür* "hump") is detectable, and a triform S>D>A emerges. A phonological feature distinguishes the nominal uses of such words from adverbial and adjectival instances: the nominal reading displays long and the adverbial/adjectival overlong[12] quantity (see Tiits 1982). The prolonged quantity may reveal the emancipation of the form as well as emphasis. Some, but not all such examples have orthographic distinction: *aukus* "hollow" pro *augus* "in the hole", *harkis* "spreaded" pro *hargis* "in the fork", *mõlkis* "dented" pro *mõlgis* "in the dent". Importantly, the static locative semantics (inessive and adessive cases) contributes to the adjective interpretation. The directional (illative/elative; allative/ablative) forms of the same words (*mõlki, mõlgist*; *laokile*, *laokilt*) can be interpreted either as adverbs or the respective case forms of nouns but not as adjectives (i.e. as a type S<>D according to our typology).

A subtype of D<>A biform uses an adjective (by its origin) as a modifier of another adjective (see 10–11). Using (emotive) adjectives as intensifiers brings forth expressivity and new biforms may emerge on that ground. These cases are handled as a pair of homonymous entries in dictionaries. As adjectives, they inflect and agree with the head noun of the phrase; as adverbs, they stay uninflected.

(10) *hirmus lugu* "dreadful affair" ~ *hirmus armas* "awfully cute"
(11) *kaunis aed* "beautiful garden" ~ *kaunis külm* "pretty cold"

1.3 D<>S [adverb<>noun]

There are basically two ways for intersection of adverbs and nouns. First, genuine modifiers are used in the position and function of nouns, occasionally. It appears that adverbs of manner are especially prone to such a shift. One subtype of D<>S biforms comprises words of foreign origin that are opaque in respect of their original meaning and are interpreted as denoting things or persons characterised by associable manners (cf. examples (12–13). A pattern of systematic polysemy, MANNER – THE PERSON/BEHAVIOUR IN THAT MANNER, appears. Another group of words reflecting a similar shift are the expressive descriptors of manner (ideophonic stems that can also be (fully or partially) reduplicated, cf. 14–15).

The second D<>S subtype comprises expressive words (often derivatives, e.g. with the diminutive suffix *-ke* as *kübeke* (see 16), *sutike, natuke, tsipake*, all meaning roughly "a tiny thing" and "a little X") that can be used to modify adjectives or adverbs. These words function as heads of a quantifying phrase (a noun) and they are subjected to declination in the same way as the numerals in the same position (*kübeke aega* [speck time-PART] "tiny bit of time"). They are interpreted as adverbs of measure (*kübeke pikem* [speck longer] "a tiny bit longer"); cf. the nominal use: *üks kübeke* [one speck] "one tiny bit". Another subtype comprises case forms of nouns functioning as modifiers and situated at diverse grammaticalization stages between noun and adverb (see 17–18). The PoS categorisation depends on the level of emancipation of the forms in these specific functions and meanings. Defining this level presumes case studies and individual examination.

(12)	D>S	*allegro*		"quickly"	"a quick and lively composition"
(13)		*inkognito*		"unrecognizably"	"appearance with hidden identity"
(14)		*jõnks*		"abruptly"	"jounce"
(15)		*liga-loga*		"not taken care of"	"trash, rubbish"
(16)	S>D	*kübe-ke*	speck-DIM	"tiny grain or speck"	"tiny bit of"
(17)		*ideaali-s*	ideal-INE	"in ideal"	"ideally"
(18)		*kahju-ks*	damage-TRA	"to damage/harm"	"unfortunately"

[12] Estonian has a three-way quantity system in disyllabic feet (Lehiste 1997; Krull & Traunmuller 2000): short (quantity 1), long (quantity 2), and overlong (quantity 3).

1.4 I<>D [interjection<>adverb]
The intersection of adverbs and interjections happens when a specific type of genuine interjections is used as an expressive modifier in the role of an adverb. The cases in our database indicate that the I<>D relation reflects the expressions of manner. The interjections involved in this type are of a kind that imitate a movement and often also a sound accompanying that movement. The expressive implications of I<>D biforms are reflected in their ideophonic phonological form (see 19–27); there are often reduplicative patterns (21–23) and these biforms may also contain specific suffixes implicating manner (*-ti/-di, -ki*, cf. 24–27). The PoS shifting from interjection to adverbs hence seem to comprise a pattern of systematic polysemy, SOUND – MANNER OF MOVEMENT, and, as such, should be represented systematically in a dictionary.

(19)	*siuh*	"noise accompanying a fast strike or friction"
(20)	*vups*	"jump out of"
(21)	*vutt-vutt*	" [child's movements] quickly, with short steps"
(22)	*sulla-sulla*	" [child's movements in water, e.g. in bath] splash, paddle"
(23)	*kippadi-kappadi*	"clip-clop; gallop, prance with clacking noise"
(24)	*klõmdi*	"bang, plump; tiff"
(25)	*müraki*	"bang"
(26)	*siuhti*	"whizz (off) "
(27)	*vupsti*	"jump out of, slip"

Another subtype of I<>D biforms comprises adverbs used as affirmative interjections: *hästi,* "well" *just* "exactly", *justament* "exactly" (in humorous register) or the other way around: affirmative interjections that can be used as adverbs, e.g. *okei* "OK"*, oolrait* "all right". The I<>D biforms can be used in the position and function of nouns, too, which leads to an emergence of a triform I<>D<>S. Such a triform occurred 74 times in the data retrieved from the morphological database (e.g. *plärts* "splash", *prõks* "crack").

1.5 N<>P [numeral<>pronoun]
This biform constitutes a closed set of word forms as both numerals and pronouns are closed word classes. In Estonian, pronouns can substitute for numerals (yielding sc. pronumerals as *mitmendik* "which part"; cf. also Section 2). The N<>P biforms are compounds of a pronoun and a numeral (see 28–29). The word forms are interpretable as numerals since they occur in quantifying constructions and they are classified as pronouns due to their semantic emptiness and deictic nature. The practical solution regarding the N<>P biforms' presentation in DicEst is to tag them with both labels. A special case of N<>P is the use of some numerals (e.g. *üks* "one") as determiners marking definiteness/indefiniteness and accompanying noun phrases to indicate that its referent is identifiable (a tendency in a language lacking grammatical articles, see e.g. Dryer 2013a–b), also Hint, Nahkola, Pajusalu 2017: 66–67).

(28)	*mõnisada*	some + hundred	"a couple of hundred"
(29)	*paarkümmend*	couple + *-teen*	"about twenty"

5.2 Unidirectional Types

2.1 V>S [verb>noun]
The double interpretations of V>S ambiforms in our data occur due to homonymy of form, unexceptionally (e.g. *mõistes* [concept-INE] and [understand-GER]). The V>S shift comprises potentially a large set of ambiforms due to the productive and regular patterns of nominalisations (the action nouns derived with the suffix *-mine* and agent nouns derived with *-ja*) applicable to the verb stems. Such nominalisations obtain all the syntactic functions of nouns. They are analysed as nouns by automatic morphological analysis and create no disambiguation problems. The regular nominalisations are presented in DicEst only occasionally, e.g. *võimlema* "to work out" > *võimlemine* "gymnastics" > *võimleja* "gymnast". The nominalisations will probably find their way into DicEst more often as there is no need to save the space in the era of electronic dictionaries and the goal is as detailed coverage of the vocabulary as possible. It would be useful to explicate the derivational link to the respective verbs while presenting them and add the regular derivational morphology to the block presenting conjugation.

2.2 V>A [verb>adjective]
The V>A biforms comprise the non-finite forms of verbs functioning as attributes – participles and supines – occurring both in verb phrases and, as modifiers, in noun phrases. Distinguishing these two types of usages is a huge problem for the automatic morphological analysis – 97% of the non-disambiguable word forms belong to this type.[13] The V>A biforms are also problematic for the lexicographers because it is not obvious when a verb form has emancipated enough to be handled as an autonomous dictionary entry rather than a regular conjugational verb form. There are examples of past participles in our database (see 30–34), present participles (35–37), and abessive supine forms expressing a

[13] The statistics originates in our analysis of the Corpus of the Estonian Web (etTenTen), containing 270 million words from 686 000 web pages, to be published.

non-performed obligatory action (see Viitso 2003: 64; examples 38–39). In some cases, the adjectival forms of verbs can be further subjected to nominalisation, in which case a triform V>A>S emerges (40–42).

(30)	*armunud*	"fallen in love"	
(31)	*joobnud*	"drunken"	
(32)	*surnud*	"dead"	
(33)	*austatud*	"honorable"	
(34)	*suletud*	"closed"	
(35)	*siduv*	"binding"	
(36)	*lööv*	"striking"	
(37)	*hävitav*	"destroying"	
(38)	*rääkimata*	"untold"	
(39)	*värvimata*	"unpainted"	
(40)	*alluv*	"subordinating"	"subordinate (person) "
(41)	*liidetav*	"adding"	"addend"
(42)	*tagaotsitav*	"wanted"	"persona non grata"

2.3 S>K [noun>adposition]

The S>K biforms represent certain forms of nouns (typically in locative or other semantic cases) that are (at least nearly) entrenched to the extent of independent lexical units. Such a shift reflects the process of grammaticalization. The meaning of the emancipated item has undergone bleaching but has not yet lost its semantic content fully (see (43–45). The word forms are accompanied by a noun in genitive case form in most of the cases (forming a head-complement relation, like adpositions). The process of evolving adpositions (and adverbs, see type 1.3, the adverbs emerging from case forms of nouns) from nouns in locative cases is in Estonian ongoing one (see e.g. Grünthal 2003: 26; EKG II: 38). The lexicographic presentation of such forms would depend on the level of their entrenchment, and the syntactic and semantic analysis of the forms in context.[14]

(43)	*aja-l*	time-ADE	"in the time of, during"
(44)	*aluse-l*	base-ADE	"on the basis of"
(45)	*andme-te-l*	data-PL-ADE	"according to"

2.4 K>D [adposition>adverb]

K>D biform is a pattern that employs a relational word either in an adpositional or adverbial function, referring often to spatial relations, for instance *juurde* "hither" and *pihta* "targeted, at". The biform is considered as syntactically dependent in the adpostional phrase, where it functions as a head of a complement (typically a noun phrase), and syntactically independent when occurring in an adverb phrase modifying a verb or an adjective. The shift from adposition to adverb can happen by skipping the nominal complement by ellipse. However, this is only one of the possible analyses, reflecting the synchronic view; diachronically, most of these ambiforms are grammaticalized forms of nouns and according to the (historical) interpretation these are tri-forms (S>D>K;[15] cf. also the D<>S subtype 1.3). Series of locative case forms (the lative, locative and separative ones, see 46–47) occur among those amibiforms. This is an instance of the case of the (prototypically) non-inflected words that some of them have (1–3) inflected forms: directional (illative or allative), static (inessive or adessive), and separative (elative or ablative) forms, depending on the verb's semantics (see Viitso 2003: 66).

(46)	*äär*	"edge"	*äär-de*	ILL	"to the edge"	*ääre-s*	INE	"at the edge"	*ääre-st*	ELA	"from the edge"
(47)	*kand*	"heel"	*kannu-le*	ALL	"to behind"	*kannu-l*	ADE	"behind"	*kannu-lt*	ABE	"from behind"

The K>D type comprises also compounds as the words with the final component *-poole* 'towards' which are exceptional[16] also because both constituents are subjected to partial case inflection (see 48–50):

(48)	*allapoole*	"downwards"	*alla* "down" + *poole* "towards"
(49)	*allpool*	"lower down"	*all* "down"' + *pool* "at about"
(50)	*altpoolt*	"from below"	*alt* "from down"+ *poolt* "from about"

2.5 V>D [verb>adverb]

The V>D biform comprises non-finite verb forms converbs, i.e. supines (e.g. the abessive supine *äraarvamata* "unbelievably"), and gerunds (e.g. *mängeldes* "easily", lit. " by playing") in adverbial function. The converbal biforms

[14] The procedures are currently under development.

[15] Habicht & Penjam (2006: 57) argue that the direction of grammaticalization in case of Estonian adpositions is generally the following: lexical form (noun+case ending) > adverb > adposition. Based on our data, we would not generalize this direction to all cases; at this point we confine ourselves to the recognition that this topic should be investigated further.

[16] Generally, only the final lexeme of a compound is subjected to declination in Estonian.

are rare in our database; because of their regularity they are included to DicEst as keywords only in the case they have adapted a deviant meaning. Another case of the V>D biform is the use of the supine in the role of modifier of the main verb, e.g. *läks minema* "went away" lit "went to go (inf)" and *tuleb tulema* "he/she leaves" lit. "he/she comes to come (inf)".

2.6 I>S [interjection>noun]

The I>S biform represents a productive pattern where prototypical (non-inflectional) interjections function as nouns. In this case, interjective word forms refer to the acts of interjecting, the activity itself would be expressed with verbal (finite) morphological forms that cannot be mixed with nominal patterns (and therefore this is not a I>S>V pattern in Estonian as it would be in e.g. English). Part of the I>S ambiforms expresses sounds of birds and animals and have ideophonic-imitative phonological form (see 51–54). Another I>S subtype are certain exclamatives, both loanwords and native words (55–57). The current lexicographic practice is to present the biforms of the interjections and respective nouns as a pair of homonyms, which is not the case, but relies on a practical principle that lexemes with different inflection would obtain individual entries The I>S biform is very regular and could be described by a pattern of systematic polysemy SOUND – THE ACT OF SOUND-MAKING.

(51) *kraaks* "croak"
(52) *nurr* "whisker"
(53) *prääk* "quak"
(54) *urr* "growl"
(55) *aamen* "amen"
(56) *braavo* "bravo"
(57) *aitäh* "thanks"

2.7 P>S [pronoun>noun]

This is a closed group as much as the class of pronouns is closed by nature. Typical P>S biform is a pronoun that has acquired a specific (conceptually richer) meaning: *mina* "I" and "self, ego"; *eikeegi* "no-one" and "person with no value". The P>S biforms tend to be presented as pairs of homonyms in dictionaries, which might not be an optimal solution because of their semantic relatedness. Incorporating the alternative interpretations as nouns into the sense menu of pronouns could be considered.

2.8 N>S [numeral>noun]

The N>S biform is a closed class as the class of numerals is closed. The numerals can be systematically used as nouns referring to the signifier of a number or a mark in a school system. Thus, a pattern of systematic polysemy NUMBER – SIGNIFIER-MARK can be postulated. Lexicographers should keep this in mind while compiling the entries for numerals. Another subtype of N>S biforms are group nouns, exploiting the numeral stems in compounds with figurative quantifiers as result (*kuradi+tosin* "the devil's dozen, 13", *must+miljon* "black million").

2.9 P>A [pronoun>adjective]

The P>A biform is a closed group as the class of pronouns is closed by nature. The word forms can be interpreted either as adjectives because they function as attributes or pronouns (in the case they are used as substituting nouns in a clause). A special PoS tag – adjectival pronoun – is coined in some dictionaries (e.g. EKSS). A typical biform comprises the root of a pronoun and an adjectival suffix (*niisugune* "such", lit. "this-like"; *samane* "same", lit. same-like; *teistsugune* "different", lit. "other-like").

2.10 J>D [conjunction>adverb]

This biform is a closed set of word forms as the class of conjunctions is closed. The intersection of adverbs and conjunctions can happen when the conjunctions are used in adverbial functions. They appear, as modifiers, typically, emphasising some constituent or a whole sentence (*ega* "nor", *justkui* "as if", *nagu* "like"). Another case of J>D biforms are the compound conjunctions that consist of a conjunction and an adverb (*niihästi ... kui (ka)* "both ... as").

6 Conclusion and Discussion

The main contribution of this study is outlining the typology of ambiforms and explaining the PoS border areas as resulting from the network-like interaction of the traditional PoS categories. Each and one of the types described above deserves a more elaborated analysis. The described network presents our current understanding based on the database of ca 3500 records; an analysis of corpora may reveal new types of connections missing from our current outline – for instance, some types described here as unidirectional may turn out to be bidirectional in nature. Therefore, we foresee that the tentative numeration presented here (reflecting the Donor-Target direction and prominence of ambiforms) could change in the later stages of the studies. Also, more specific subtypes could be distinguished and described in the future. One of the general observations arising from the descriptions above is that the noun has a special position among the interacting PoS categories. On the one hand, we found it "feeding" the syntactically dependent PoS categories as a Donor. A reason for this are the ongoing processes of grammaticalization utilising the means of nominal morphology. This

finding is in line with the study of Karelson (2005) who also addressed several nominal pairs (noun-adjective, adjective-proper noun, noun-adverb, interjection-noun) and focused separately on the phenomenon characteristic to the Estonian language, i.e. the continual supplementation of adverbial and adpositional classes by (typically locative) case forms of nouns. The phenomenon was seen as one of the biggest challenges also in the analysis of Habicht et al. (2011) concerning the problems arising in connection with PoS annotation of the corpus of Old Written Estonian. The noun-centred pairs mentioned in their study were (noun-adjective, noun-adposition, noun-adverb).

The finding in the present study is that the noun is also a popular Target: the categories with lower syntactic/semantic status could be "upgraded" to nouns while there was a need to use them in roles of subject, object or predicative. Subjecting the uninflected words occasionally to case inflection is also a possibility of morphologically rich language. Another aspect of the noun-centeredness is that in numerous cases we admitted the emergence of triforms by accepting an alternative extra analysis of the ambiform as a noun – either by nominalisation (e.g. I<>D<>S, V>A>S) or interpreting the form as a certain semantic case form (e.g. S>D>A).

Adverb occurred as the second attractive interacting PoS category. The analysis of bi-forms revealed that adverb emerged mostly as the Target (see Figure 1) with two exceptions (adverb>adjective and adverb>noun). Three adverbial pairs (adjective-adverb, interjection-adverb, adposition-adverb) are described also by Karelson (2005); we were able to introduce two additional pairs (verb>adverb and conjunction>adverb).

The interactions of the adjective (with noun, verb, and adverb) have also been described in the previous studies (Karelson 2005; Habicht et al. 2011; Paulsen et al. 2019). We were able to explicate the major role of the A<>S biform (40–50%) in the pool of ambiforms (see Table 2). This finding repeated the result of the previous metalexicographic study where lexicographers mentioned the noun-adjective ambiform as the most problematic area of PoS tagging (Paulsen et al. 2019). The adjective-noun conversion in Estonian is neither a clear case of the notion of syntagmatic category mixing (the syntax and semantics of one class are mixed with the morphological properties of another class) nor paradigmatic category mixing (a word has morphological properties of two categories, see Nikolaeva & Spencer 2019: 42), and the deadjectival person noun can have both a generic and referential interpretation.

A striking result of the ambiform typology is the weighty role of expressive vocabulary (onomatopoetic and descriptive words) in PoS ambiguity overall, explainable with the richness of ideophonic language in Estonian. Ambiguity with respect to lexical categorisation is a characteristic feature of the descriptive words in particular: as the results of Mikone's (2002: 154) study show, "there is an adjective element in descriptive substantives and an adverbial element in descriptive verbs". Another point of view on the categorisation of ideophonic expressions, traditionally also involving interjections, in Balto-Finnic (including Finnish and Estonian, among other) languages is to treat these words as a special class of ideophonic verbs, substantives, and particles (Mikone 2001: 225). We observed that the expressive/emphatic function might be a reason also for the phonological change of some ambiforms while shifting from inflected forms of nouns into the uninflected class of adverbs and atypical uninflected adjectives. The strive for increased expressivity was suggested also as a cause for certain adjectives to turn into intensifying adverbs and for a minor subtype of figurative quantifiers to occur.

Systematic polysemy was found as a predictive force behind the emergence of ambiforms. Its role in lexicology and lexicography has been described mostly from the perspective of sense alterations (Langemets 2010), its PoS altering potential is not fully discovered, yet. Surprisingly, the typical word-class altering device in Estonian, suffixal derivation (see, e.g. Vare 2006: 199) did not appear to be particularly problematic (e.g. nominalisations); neither did adjectival uses of verbal forms (participles, infinitives) – which complicate automatic PoS tagging of text corpora – emerge as problematic in the morphological database nor the metalexicographical study (Paulsen et al. 2019). This may be due to the unsteady status of these forms or even the general principle of exclusion of more or less regular phenomena from the (paper) dictionaries.

Lexicographers strive to order the lexemes correctly. The typology of ambiforms presented in this study would hopefully help the lexicographers to raise their awareness of the potential alternative interpretations. Besides, the knowledge about the proneness of certain PoS to combine with others either in bi- or unidirectional manner can be built directly into the dictionary writing system, which, then, would guide the lexicographer to check for the most probable alternative uses (and categorisations) of the headword. Awareness of the role of expressivity as a potential force bringing forth PoS ambiguity can be put to work in lexicographic work, too. It can be done, for example, by creating a module of the lexicographers' tool that reminds them to check for alternative uses of the word whenever the PoS label "interjection" has been entered. Knowledge about productive patterns like systematic polysemy is useful in lexicographic work as it facilitates cross-checking of the meanings (and PoS categories) involved. An understanding of lexical categorisation benefits of the awareness of border areas, which, at bottom, reflects the essence of language.

7 References

Blensenius, K., von Martens, M. (2019). Improving Dictionaries by Measuring Atypical Relative Word-form Frequencies. *Proceedings of eLex 2019 conference*. 1−3 October 2019. Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o., pp. 660–675.

Chomsky, N. (1974). *The Amherst Lectures. Lectures given at the 1974 Linguistic Institute*, University of Massachusetts, Amherst; Université de Paris VII.

CombiDic = The Combined Dictionary of Estonian (2020). I. Hein, J. Kallas, O. Kiisla, K. Koppel, M. Langemets, T. Leemets, M. Melts, S. Mäearu, T. Paet, P. Päll, M. Raadik, M. Tiits, K. Tsepelina, M. Tuulik, U. Uibo, T. Valdre, Ü. Viks & P. Voll (eds.). Eesti Keele Instituut. Sõnaveeb 2020. Accessed at: https://sonaveeb.ee [14.2.2020].

Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Culicover, P. W. (1999). *Syntactic Nuts*. Oxford University Press: Oxford.

Dryer, M. S. (2013a). Indefinite articles. In S. Matthew Dryer, M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Accessed at: http://wals.info/chapter/38[14.2.2020].

Dryer, M. S. (2013b). Definite Articles. In S. Matthew Dryer, M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Accessed at: http://wals.info/chapter/37 [14.2.2020].

ECD = Eesti keele naabersõnad [The Estonian Collocations Dictionary]. (2019). Kallas, J., Koppel, K., Paulsen G. & Tuulik, M., Institute of the Estonian Language. Accessed at: http://www.sonaveeb.ee. [14.2.2020].

EKG I = Erelt, M.,Kasik, R., Metslang, H., Rajandi, H., Ross, K., Saari, H., Tael, K. & Vare, S. (1995). *Eesti keele grammatika I. Morfoloogia*. Sõnamoodustus. Tallinn: Eesti TA Eesti Keele Instituut.

EKG II = Erelt, M., Kasik, R., Metslang, H., Rajandi, H.,; Ross, K., Saari, H., Tael, K. & Vare, S. (1993). *Eesti keele grammatika II. Süntaks.* Lisa: kiri. Tallinn: ETA Keele ja kirjanduse instituut.

EKSS = Eesti keele seletav sõnaraamat I–VI [The Explanatory Dictionary of Estonian]. (2009). M. Langemets, M. Tiits, T. Valdre, L. Veskis, Ü. Viks, P. Voll (eds.). Institute of the Estonian Language. Tallinn: Eesti Keele Sihtasutus. Accessed at: http://www.eki.ee/dict/ekss/ [14.2.2020].

ES2019 = Eesti keele sõnaraamat [The Dictionary of Estonian]. (2019). M. Langemets, M. Tiits, U. Uibo, T. Valdre & P. Voll, (eds.); Institute of the Estonian Language. Accessed at: http://www.sonaveeb.ee. [14.2.2020].

Grünthal, R. (2003). *Finnic Adpositions and Cases in Change*. Suomalais-Ugrilaisen Seuran toimituksia 244. Helsinki: Finno-Ugrian Society.

Erelt, M. (2017). Sissejuhatus süntaksisse. In M. Erelt, H. Metslang (eds.) Eesti keele süntaks. Tartu: Tartu Ülikooli Kirjastus, pp. 537–564.

Habicht, K., Penjam, P. (2006). Kaassõna keeleuurija ja -kasutaja käsituses [Adpositions as viewed by a linguist and by a language user]. Emakeele Seltsi aastaraamat 52. Tallinn, pp. 51–68.

Habicht, K., Penjam, P., Prillop, K. (2011). Sõnaliik kui rakenduslik probleem: sõnaliikide märgendamine vana kirjakeele korpuses [‘Parts of speech as a functional and linguistic problem: annotation of parts of speech in the corpus of Old Written Estonian’]. *Estonian Papers in Applied Linguistics*, 7, pp. 19–41. Accessed at: https://doi.org/10.5128/ERYa7.02 [14.2.2020].

Hint, H., Nahkola, T., Pajusalu, R. (2017). With or without articles? A comparison of article-like determiners in Estonian and Finnish. Lähivõrdlusi. *Lähivertailuja*, 27, pp. 65−106.

Karelson, R. (2005). Taas probleemidest sõnaliigi määramisel [Once more on the issues of determining parts of speech]. *Estonian Papers in Applied Linguistics*, 1, pp. 53– 70.

Kasik, R. (2015). Sõnamoodustus [Estonian word-formation]. Eesti keele varamu I. Tartu: Tartu Ülikooli kirjastus.

Koppel, K. (2020). Näitelausete korpuspõhine automaattuvastus eesti keele õppesõnastikele [Corpus-Based Automatic Detection of Example Sentences for Dictionaries for Estonian Learners]. PhD thesis. Tartu: Tartu Ülikooli Kirjastus.

Koppel, K., Tavast, A., Langemets, M. & Kallas, J. (2019). Aggregating dictionaries into the language portal Sõnaveeb: issues with and without a solution. In I. Kosem, Z. Kuhn, T. Correia, M. Ferreria, J. P. Jansen, M. Pereira, J. Kallas, M. Jakubíček, S. Krek & C. Tiberius (eds.). *Proceedings of the eLex 2019 conference. 1−3 October 2019*, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o., pp. 434−452.

Krull, D., Traunmüller, H. (2000). Perception of quantity in Estonian. *Proceedings of fonetik* 2000, pp. 85–88.

Langemets, M. (2010). Nimisõna süstemaatiline polüseemia eesti keeles ja selle esitus keelevaras [Systematic polysemy of nouns in Estonian and its lexicographic treatment in Estonian language resources]. PhD thesis. Tallinn: Eesti Keele Sihtasutus.

Langemets, M., Uibo, U., Tiits, M., Valdre, T. & Voll, P. (2018). Eesti keel uues kuues. Eesti keele sõnaraamat 2018 [Estonian lexis revisited: The Dictionary of Estonian 2018]. Keel ja Kirjandus, 12, pp. 942−958.

Lehiste, I. (1997). Search for phonetic correlates in Estonian prosody. In I. Lehiste, J. Ross (eds.) *Estonian prosody: papers from a symposium*. Tallinn: Institute of Estonian Language, pp. 11–35.

MAB = Eesti Keele Instituudi eesti keele morfoloogiline andmebaas (2019). Ü. Viks, I. Hein, & K. Tsepelina (Koost). Eesti Keele Instituut. Sõnaveeb. Accessed at: https://sonaveeb.ee [14.02.2019].

de Marneffe, M.-C., Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J. & Manning C. D. (2014). Universal Stanford Dependencies: A cross-linguistic typology. In: *Proceedings of LREC*.

Mikone, E. (2001). Ideophones in the Balto-Finnic Languages. In F. K. Erhard Voeltz & C. Kilian-Hatz (eds.) Ideophones. Amsterdam: John Benjamins, pp. 223−233.

Mikone, E. (2002). *Deskriptiiviset sanat: määritelmät, muoto ja merkitys* [Descriptive words: definitions, form and meaning]. Helsinki: SKS.

Nikolaeva, I., Spencer, A. (2019). *Mixed Categories. The Morphosyntax of Noun Modifications*. Cambridge University Press.

Paulsen, G., Vainik, E., Tuulik, M. & Lohk, A. (2019). The lexicographer’s voice: word classes in the digital era. *Proceedings of eLex 2019 conference*. 1−3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o., pp. 319–337.

Paulsen, G., Vainik, E. & Tuulik, M. (2020). Sõnaliik leksikograafi töölaual: uuring sõnaliikide rollist tänapäeva leksikograafias [On word classes in contemporary lexicography. The lexicographers’ view]. *Estonian Papers in Applied Linguistics*, 16.

Pedersen, B. S., McCrae, J., Tiberius, C. & Krek, S. (2018). ELEXIS – a European infrastructure fostering cooperation and information exchange among lexicographical research communities. In *Proceedings of GlobalWordNet Conference 2018*. Singapore.

Quecedo, J. M. H. (2019). Neural models for unsupervised disambiguation in morphologically rich languages. Master Thesis in the University of Helsinki.

Rosch, E. (1978). Principles of categorization. In Cognition and categorization (27–48), E. Rosch & B. B. Lloyd (eds.). Hillsdale, Lawrence Erlbaum, New York.

Smith, M. C. (2015). Word categories. In J. R. Taylor (ed.) The Oxford Handbook of the Word. OUP Oxford: Kindle Edition.

Tavast, A., Langemets, M., Kallas, J., Koppel, K. (2018). Unified Data Modelling for Presenting Lexical Data: The Case of EKILEX. *Proceedings of the XVIII EURALEX International Congress: EURALEX: Lexicography in Global Contexts, 17−21 July 2018, Ljubljana*. J. Čibej, V. Gorjanc, I. Kosem, S. Krek, (eds.). Ljubljana University Press, Faculty of Arts, pp. 749−761.

Taylor, J. R. (2012). *The Mental Corpus: How Language is Represented in the Mind*. Oxford: Oxford University Press.

Tiits, M. (1982). Seisundiadverbidest [On state adverbs]. Keel ja Kirjandus 1, pp. 17–21.

Vare, S. (2006). Adjektiivide substantivatsioonist ühe tähendusrühma näitel. [On substantivisation of adjectives: Analysing a semantic group] E. Niit. Keele ehe. Tartu: Tartu Ülikool. Tartu Ülikooli eesti keele õppetooli toimetised; 30, pp. 205−222.

Viitso, T.-R. (2003). Structure of the Estonian language: Phonology, morphology and word formation. In M. Erelt (ed.) Estonian language. Tallinn: Estonian Academy Publishers, pp. 1−9.

Acknowledgements

This work was supported by the Estonian Research Council grant PSG227.

Abbreviations: ABE = abessive; ADE = adessive; ALL = allative case; DIM = diminutive; COM = comitative; COMP = comparative; ELA = elative; GEN = genitive case; GER = gerund; ILL = illative case; INE = inessive; PART = partitive case; PTC = participle; PL = plural; SUP = supine; TERM = terminative; TRA = translative.