COMPILATION OF THE DATABASE OF ESTONIAN PHRASES

Katre Õim, Asta Õim, Anne Hussar, Anneli Baran

The earliest entries of the *kõnekäänud* '(proverbial) phrase' files in the Estonian Folklore Archives presumably originate in the 17th and 18th century; most of the phrases have been collected in the late 19th and 20th century. The material has been copied from the Estonian Folklore Archives as well as the manuscript catalogues of the Estonian Cultural History Archives, the dialectal archives of the Estonian Language Institute and the Estonian Language Society, and also from various printed publications. In 1994 the estimated number of archival index cards was *ca* 200,000 (Baran 1999: 2; Õim 1997: 3), whereas presently the number remains somewhere below 150,000, and the total number of expressions is about 160,000.

The accumulation of the material over a long period, the uneven quality of entries, the fact that proverbial expressions are subject to improvisation, etc. create various problems at working with the files. The files include several texts of dubious authenticity. Similarly to the files of proverbs and riddles, the phrase entries often include duplicates from printed sources and earlier manuscripts, revisions made by collectors and even downright fabrications. A more important problem, however, is the issue of categorising figures of speech among proverbial phrases, i.e. problems related to distinguishing between variable and invariable compounds (cf. *aher naine* 'barren woman' – variable metaphoric compound, *kadunud poeg* 'the lost son' – proverbial phrase),[1] but also determining the obligatory/invariable and facultative/variable constituents of the phrase. The categorisation of phrases, therefore, must be based on the revision and systematisation of archival material. An important tool for this work will be the constructed database of phrases, which main aim and also advantage will be fast search options and browsing in the database material.

1. ESTONIAN *KÕNEKÄÄND* 'PHRASE'

1.1. *Kõnekäänd* as a word, phrase, sentence

(Proverbial) phrase is very heterogeneous in that it encompasses phraseology in the strictest sense (in borderline cases compound verbs, compound words, etc.) but also longer figurative expressions (dialogues consisting of several sentences, etc.). Some of these are related to the structure and lexica of language, i.e. they are purely linguistic invariable compounds, some exist as speech-bound folkloric texts, i.e. they are figures of speech. Hence the problems with finding a uniform criterion, both linguistically and theoretically-methodologically, not to mention the theories of trope structure, in the study of sayings.

For a satisfactory result in categorising the object of research, it is important to know it and to be able to determine it. The cognitive value of any categorisation depends on the significance of information about the object and on how the categorisation characteristics describe it. Which characteristics should the classification be based on? Paremiologists, with a tongue in cheek, say that a phrase is something that is neither a proverb nor a riddle. In other words, *kõnekäänd* is a figure of speech that does not generalise the object (like proverb), nor is it an allegorical hidden description (like riddle). Obviously, this definition is not sufficient to explain the term (more on this issue, see: Krikmann 1997: 52–74, 151–171). The distinction between sayings and other figurative means of expression, say, linguistic metaphors, is the classification of figures of speech on the basis of certain characteristics.

Language, a mechanism transforming thought into text, can function only through the means of certain semantic units. In the text, each linguistic unit can be used only in a certain semantic context, syntactic position and in certain morphological forms. In its communicative function primary significance is attached to the sentence, not any smaller linguistic units. When consisting of several sentences, a saying functions as a syntactical constructor of a sentence: as a phrase, a noun phrase (*soss-sepp* 'bungler', *kahe pere koer* 'a dog of two families', 'a hypocrite', *kadunud poeg* 'the lost son'), a verb phrase (*kintsu kaapima* 'to bow and scrape', *kannab*

keelt 'blows whistle', *pistis punuma* 'scampered away') or a compound phrase (*nael sülti* 'like a pound of jelly' – 'an effeminate (man)', 'a sissy', *viies vesi taari peal* 'the fifth water on kvass' – 'a very distant relative'), etc. *Kõnekäänd* as a phrase has an exocentric structure, i.e. it is not equivalent with the basis nor the extension of the phrase; the syntactic and the semantic function is carried by the entire phrase. The components of the phrase cannot be omitted from the sentence (from the (proverbial) phrase) without a change in the contents and/or syntactic structure of the sentence (From the (proverbial) phrase) (cf. *Ta on kahe pere koer* 'He is "a dog of two families"' (i.e. a hypocrite). **Ta on koer* *'He is a dog' **Ta on kahe pere* *'He is of two families'), while substitution is possible (cf. *Ta on kahe pere koer* 'He is a dog of two families', *Ta on mitme pere koer* 'He is a dog of several families', *Ta on kahe talu koer* 'He is a dog of two farms', *Ta on mitme talu koer* 'He is a dog of several farms').

Due to the versatility of phrase forms this definition (*kõnekäänd* as a phrase) can be applied only to the core of the genre. The peripheral part of sayings, i.e. various speech formulae, wishes, greetings, comments of situations, etc. are incomplete, elliptical and complete sentences or units consisting of several sentences. Proverbial phrase as a typological category, therefore, consists of units of different levels: on the one hand phrases that share characteristics with words, and on the other hand sentences with characteristics of a text. Or, if we use the terminology of Arvo Krikmann: *kõnekäänd* is sometimes a figurative "building material", sometimes a "work" (Krikmann 1997: 53). But as is often the case with other classifications, the taxonomy of sayings leaves room for ambiguities. The problem does not necessarily lie in the inadequacy of categorisation criteria. If we do not regard language as a static system but rather as the result of a historical process and as the starting position of further progress, we will once again realise that ambiguity in language is quite natural.

It is realistic to view any classes of language units, incl. sayings, as infinite classes. They contain elements that belong undoubtedly to the same class as well as those that rather belong to another. This is often the case with the border between the categories of sayings and proverbs, and between sayings and linguistic metaphors. Sometimes it is fairly difficult to decide whether it is a generalisation, a

regularity, a formulation of a norm (i.e. proverb) or a situation, a figurative characterisation of an irregularity (i.e. proverbial phrase).

Sayings, particularly the units lesser than the sentence are prone to change and transformation. At times they assume different forms in language and, similarly to word, they are easily derived, borrowed, assimilated, overused, forgotten and rediscovered, i.e. they function as language units rather than folkloric texts. Ordinarily, this is not the case with proverbs and riddles.

1.2. *Kõnekäänd* as the characterisation of an irregularity

On the content level we should look for the characteristics of phrase as a class in the sentence. In the semantic aspect a sentence designates a situation. The constructor of situation type sentences is called semantic predicate, the participants in the situation – actants (Erelt & Kasik & Metslang et al. 1993: 11 ff). These semantic functions can also be attributed to sayings, in case we regard it as the constructor of a sentence, governing the characterisation of a given situation as an irregularity. This is, in fact, the most general semantic characteristic of phrases. This characteristic, unfortunately, is far too general for defining the term, so we have to turn our attention to pragmatics. In the pragmatic sense a sentence is a message, something is known about something else. The part of a sentence, which communicative function is to be the starting point of the sentence is called theme, the rest of the sentence is called rheme. A saying with its communicative function to inform of the qualitative and quantitative characteristics, special and temporal features, modality, degree, locality, cause, etc. is clearly rhematic. In nominative function a phrase which names the object based on its characteristics, is generally expressive and conveys evaluation.

Due to the exocentric structure of a phrase, its meaning is formed as a result of the cooperation of syntactic structure, semantic functions and semantic transformation. In the Estonian language, for instance, the word combination *kärnane lammas* 'scabby sheep' does not stand for an ordinary person, but for someone who's "morally corrupt". Its specific meaning "unusual in the physical sense, unlike others", i.e. scabby, mangy, has been transformed abstract; the semantic transformation is governed by an attribute. The struc-

tural component *'lammas'* (sheep) is not associated with a person as an ordinary prototype ("a bashful animal"), but materialises the stereotype "man is animal" and the meaning is determined by the semantically transformed attribute of the content word. This could also be explained from the grammatical perspective, since the function of the attributive complement component of the construction is the in-depth explanation of the meaning.

If a saying follows a sentence structure, then its characteristics should be searched from the aspect of communicative-modal meaning. The pragmatic meaning of complete, incomplete and elliptical sentences is formulated in grammar. The communicative-modal meanings of a sentence are statement, question, order, wish and exclamation. These are expressed with sentences of statement (*Ahi lagunes ära* 'The stove is broken.'), interrogative sentences (*Kas tunned tukivingu?* 'Can you smell the smoke?'), imperative sentences (*Sõida seenele!* 'Go to hell!'), exclamatory sentences (*Püha jumal!*'Oh, god!', *Oh sa poiss!*'Oh, boy!') and sentences expressing wish (*Et see sulle kurku kinni jääks!* 'That you would choke on it!'). Phrasal sentences (*kõnekäänulised laused*) are typical in that they often perform the secondary rather than the primary communicative-modal function of the sentence. E.g. the question in *Kas sa klaassepa poeg oled?* 'Are you the son of a glazier?' covers several communicative-modal stages in the Estonian language: a question; an order of dissapproval, so that the person would move out of the way; a statement that somebody's blocking the light or view. Phrasal questions are usually not intended to elicit a reply, they are rhetorical questions. Imperative statements may express neutral, approving, pejorative or slightly disapproving attitude (*Lase käia!* 'Go on!', *Mine metsa!* 'Go to hell!', *Olgu olla!* 'Make it sure!'). A jussive command (*Sadagu või pussnuge!* 'Even if it rains cats and dogs!') expresses an obligatory statement that in spite of all, you'll have to go. The main purpose of the wishful sentence is to indicate that the speaker finds the performing of the event necessary and positive, but does not seek to actually perform it. The wish expresses the attitude towards the event, which according to the speaker may but does not necessarily have to happen (Erelt & Kasik & Metslang et al. 1993: 178). Phrasal sentences expressing wish are complete imperative sentences (*Olgu muld sulle kerge!* 'May you rest in peace!'), incomplete sentences without a verb (*Jõudu tööle!* 'Good luck!' *Jätku*

leiba ~ leivale! 'May your bread last' – 'Bon appetit!'), or elliptical imperative sentences (*Pipart sulle keele peale!* 'Pepper on you tongue!' i.e. 'Touch wood!' 'Knock on wood!'). Negative attitude is often formulated as an unreal wish (*Et vihmavari su kõhus lahti läheks!* 'That an umbrella would open in your stomach!'). Exclamatory sentences indicate that the speaker is surprised and taken aback. Unlike the wish, the meaning of exclamation does not contain the evaluation of necessity or reality, but expresses an emotive evaluation on something that has already been given the evaluation of reality (Erelt & Kasik & Metslang et al. 1993: 179). The constituents of an exclamatory phrase have become desemanticised (*Põrgu päralt!* 'For the hell with it!', 'What the hell?' *Sa sinine sitikas!* 'You blue bug!', 'You bugger!' *Välk ja pauk!*, 'Bang and thunder!', 'Confound it!', *Sa heldene aeg!* 'Oh, dear!').

The meaning of phrasal sentences, therefore, depends largely on the secondary communicative-modal functions of a given sentence. One ways to find a saying of a certain meaning in a corpus of phrases is to designate phrasal sentences by communicative purposes, such as statement and evaluation. Although phrases include all five communicative sentence types, their communicative purpose is to make a statement or express attitude towards something, not to inquire or order.

Presently, however, phrases cannot be searched by their semantic meaning. During the typological process we have attempted to assemble expressions with closer semantic meaning, though presently the search by topics is possible only through a cross-type reference system (*Aabramit näinud* 'drunk' – see *Aabramiga juttu ajanud* 'drunk'). The entire files, however, does not include this information. Single words or word combinations can be found only if they are entered as headwords.

2. ORGANISATION OF THE FILES

Before starting working at any larger body of material, the material has to be first systematised. This is also the case in turning any extensive files into a database. The first step is to define the material, then form a structure for the database, and so on.

In the 1930s Rudolf Põldmäe initiated the compiling of the thematic files of the Estonian phrases, which was supposed to be analogous to the thematic files of the Estonian jokes. Unfortunately the project came to a halt during the world war. Even though the typological work at the files of Estonian phrases was not restarted before 1994, the files were still quite actively used. The files were used for different purposes by scholars and other people interested, although the phrases were distinguished only by places of origin and searching within the material by lexical, semantic or other features was extremely laborious. In the 1950s and 1960s the Estonian phrases were studied mainly by Ingrid Sarv, who has published several articles and maintained a doctoral thesis "On the Subcategories and Functions of Estonian Phrases" (1964). Feliks Vakk, a scholar of phraseology, and others have also made several references to the files in their research.

2.1. Authentication of phrases

The Estonian phrases included in the files are, naturally, not of purely Estonian origin. International phrases can be found in the traditional as well as non-traditional material from published sources, such as utterances of literary origin, biblical expressions, translations. Also, we should keep in mind that all this material originates in the same area of spread and usage – Estonian dialects, the oral or the written Estonian language.

Distinguishing authentic phrases from inauthentic ones has been facilitated by our previous experience in determining the authenticity of proverbs and riddles and by the additional files compiled in the process. What we mean by the authentication process is determining the ethnic authenticity of phrases as well as revising and supplementing the original data of entries (collector, parish, year of collection, etc.).

Unlike working with riddles and proverbs, the work with phrases is mostly based on the electronic database. The compilation stories of the academic publications of proverbs, riddles and phrases form a chronological line, in a sense, stretching from paper and pencil to the electronic means. Parallelly to the typological work of proverbs and riddles, archival data was added to the entries and the ethnic

authenticity of the material was determined (this was the most time-consuming part of the preparatory stage of the publication).*Estonian Proverbs* ("Eesti vanasõnad") was based on a manuscript draft and a typewritten copy, which were twice revised during the printing process. The organisation of the riddle files was extremely thorough and its compilers may pride themselves on having touched nearly all the correspondence materials that include riddles. (In fact, sacrificing the eyesight of the compilers, and in order to protect the catalogues of Eisen and Hurt, microfilms were used.) Only after the files appeared to be organised the material was entered into the computer. The printed publication of the Estonian riddles *Eesti mõistatused*, was based on the same electronic database.

The starting point of the database of Estonian phrases is principally different. After the initial typological work on the files, it was decided to enter the information included to the database and only then determine the authenticity of the phrases and revising and supplementing the original data, employing the electronic means and solutions. Hopefully, it will speed up the entire work process and will help to avoid smaller errors. The search engine enables widely different searches: the material of a collector or a correspondence, for example, can be easily traced in the entire corpus, and, if necessary, information on the collector; also, the year of collecting or geographical origin can be revised or specified. Analogous check of possible errors was also conducted on the database of riddles (Hussar & Krikmann & Saukas & Voolaid 2001: 11).

2.2. Insufficient collecting data

People organising and supervising the collection of folklore, from Jakob Hurt to the present researchers at the Estonian Folklore Archives, have always underlined the importance of writing down merely the oral tradition, and not include any embellishments, improvements or added information from other sources by the collector (see e.g. Krikmann 2001: 29 ff).

The original card files, which were used for working with all these three folklore genres, have been compiled as a result of the work by numerous copyists during a long period of time, much like the Folklore Archives itself. Requirements for copying have changed in

times, and that explains the uneven quality and missing data of copied cards. Sometimes, information on original sources is altogether missing. The material has arrived at the archives in different ways: the archival sources include written drafts, copied passages from books for personal use, and other sources that are clearly not intended for preservation. Material has been found even from single sheets with no information of its origin whatsoever. Regardless of all our efforts a part of this material remains unidentified: the collector will be anonymous, the time and place of collection unknown. Some of the inaccuracies in the riddle files could be corrected as a result of finding out the names of actual collectors. Some files, for instance, include expressions collected by school pupils, whereas the name that appears on the card is that of their teacher, who had sent the material to the archives. The same applies to material collected on expeditions. In some cases the names of actual collectors are revealed only in the foreword or preface of the sent material, where the sender has mentioned his/her assistants. After the original collectors have been determined, the mistakes have been corrected.

As to the data about collection place, we can also rely on previous work. If it has been decided, for example, that the proverb and riddle material sent by J. Sandra originates most likely in Vastseliina or Setu, and the material by E. Poom on the border of Rapla and Märjamaa parish, then further checking is unnecessary. The origin of folkloric material is usually determined to the accuracy of parish. The constant reforming of administrative units has moved the borders of parishes, and in cases where a village is mentioned as the place of collection, it is quite difficult to determine which parish did it actually belong to. In some files it cannot be traced from the registered information: S. Karu (RKM II 188, 37), for example, comments the sent material as "heard from the Virumaa, Tartu and Läänemaa County". It is inevitable that some material, especially the earlier recordings, lack precise data on the time of collection. It would definitely be more informative for a database user if material without precise dating was divided analogously to the database of riddles: period 1. Before 1876; period 2. 1877–1917; period 3. 1918–1940; and period 4. 1941 and later. Making this amendment electronically would also not be too time-consuming.

2.3. Duplicates

One of the main objectives of the authentication process is to single out fabrications and copies that duplicate a manuscript or a printed source. The correspondence sent to the archives often include both proverbs and phrases: if this is the case then the characteristic data and authenticity of the given material has already been determined in the course of the systematisation of proverbs. This also saves from thumbing through older original files. Moreover, in the 1970s–1980s the material sent by known "fabricators" (August Krikmann, J. Sõggel, H. Leoke) was deleted from the files. Here are some examples of "the fabricated phrases" by J. Sõggel: *vanatüdruk kõmab nagu õõnes mesipuu* 'The old maid rattles like a hollow beehive'; *kerge nagu kelk lume peal* 'light as a sledge on the snow'; *kisendab nagu karupoeg* 'yells like a bear cub'; *pea kõigub nagu viie vedru peal* 'head rocks as if on five springs'. Fortunately, there are considerably less printed sources containing the whole series of phrases than there are those containing proverbs or riddles. While working through publications all those including phrases were first marked, then the phrases were copied. One possible explanation for the duplications is that some collectors have sent their material both to Hurt and to Eisen, later also to the Literary Museum as well as to the files of Estonian Language Society. For a researcher, recurring material based on the same manuscript is extremely troublesome. But even reliable collectors, especially when they reach a more advanced age, and no longer have the stamina to make longer collection expeditions, tend to resubmit their earlier material. Also, many draft manuscripts have been brought to the archives after the collector's death by their heritors or someone who has found them by accident.

Through the mediation of third persons the archives have received even the material written down for the collectors' own amusement. While working with riddles we were surprised to discover that even proverb collectors of impeccable reputation have slipped. At the same time several authors of inauthentic proverbs (J. Kuldkepp, T. Saks, K. J. Haus, etc.) have sent fine riddles representing local tradition. Some correspondents, however, have claimed incapable of changing their "collection methods" (A. Hiemägi, H. Karro, J. Sandra, A. Suurkask).

The authentication of both proverbs and riddles is based on the typological arrangement, i.e. we proceed from the concept of type as an entirety. It is believed that the same method can be used for determining the authenticity of phrases. For example, how to combine the phrases translated into the South- Estonian dialect by D. Lepson with their original counterparts from the publications by Eisen? There is no reason to disregard all the phrases written down by D. Lepson, as there are some that the collector has remembered originating from the authentic South-East Estonian tradition. The authenticity of a phrase type can be assessed only in this broader view: when all the expressions of the same type have been assembled, it becomes clear whose material should be compared. The same problem arises with some other collectors, who have slightly altered some expressions during copying. It is a common knowledge that F.J.Wiedemann had worked through all the earlier fiction and folklore files for his *Aus dem inneren und äusseren Leben der Ehsten* (1876), but had correspondents that have remained anonymous till the present day. In such cases we have regards the source mentioned in Wiedemann's work as authentic.

As to phrases, omitting the authentication phase has also been a topic of consideration. We do not consider necessary to review all the correspondence that include phrases, like it is done with riddles. The issue of ethnic authenticity with phrases copied from the dialectal archives of the Estonian Language Instutute and the Estonian Language Society is less problematic, since the collectors have been renowned linguists and dialect researchers. At times it seems that the collectors have been too attached to the Wiedemann's dictionary, or, to the contrary, have interpreted it too loosely. E.g. the expression *see jutt ei anna öömaja* 'this won't give you a night's lodging' is undoubtedly traditional in Lüganuse, though the number of times it was recorded on a field trip (S. Tanning, 1943 – 4; M. Must, 1949 – 2) definitely raises questions. Its traditionality is confirmed by an earlier recording from 1939 by H. Reitsnik. At the same time the entries of the dialect sector of the Estonian Language Institute may constitute a "solid" type with observable variability as well as local distribution and doubtless authenticity. Let's give some example phrases. The tradition of visiting the lasses was called: "poisi käüvä jõõsa pääl" (I. Kilk, Plv, 1925), "jõõsa pääl käümä"

(L. Unt, Plv, 1929), "poisi läävä jõõsa pale" (H. Keem, Plv, 1937). When something is completed swiftly and with determination then you can say: *ühe joonega valmis tehtud* 'done all at once, without a break' (A. Anni, KJn, 1922), *sööme ühe joonega ära* 'we will eat it all at once' (T. Kaljo, Jäm, 1925/6), *lätt üte joonega* 'went [was done] without a break' (H. Kasvandik, Plv, 1922/3), *läksi joonelt* 'went [was done] without a break' (E. Küttim, Rei, 1977). Both types occur only in dialectal texts. Or, consider the following example: *lammas läheb vähi käest villa saama* 'The sheep goes to crayfish for wool', i.e. 'will get nothing' (cf. EV 14450), which originates in *Meelejahutaja* (1885), where K. J. Haus and J. A. Veltmann have copied expressions from.This type includes phrases from the publication by Eisen, also from D. Lepson, but in the view of his inaccuracy these records cannot be considered authentic. J. Gutves from Rõuge has sent materials copied from Stein (1875). In Eisen's work the phrase *jumala viljast väsinud* 'tired of God's creation' is attributed to J. Hünerson from Karksi (Eisen 1913: 65). The general picture of the type is confirmed by the records of M. Sarv (1938) and S. Tanning (1938, 1939), also from Karksi. The inauthentic texts of this type are the texts by D. Lepson, arguably from Räpina, which originate in Eisen's material. This gives us the solid Karksi type.

A separate group of phrase types are those that originate in"the old sources" (Helle 1732; Hupel 1780). The type *laiska petma* 'to fool an idle' contains a formula *ma püüan laiska petta* 'I will try to fool an idle'*, ei lase laiska ennast petta* 'I will not let an idle to fool me', originating from A. Thor Helle, which is the only reliable text for this type. There are other printed sources, of course, but these are also traceable back to Helle, as well as six manuscriptal texts, which are traceable to Helle through Hupel or Wiedemann. The authentication process should therefore take relatively little time in comparison with the authentication of other short forms. Moreover, the number of large and old types that appear in many printed sources is relatively small.

3. TYPOLOGY OF PHRASES

In 1994 the grant project *The typology and systematisation of Estonian phrases* of the Estonian Science Foundation was launched by

Arvo Krikmann, the grant holder, and the original executors Anneli Baran, Tiit Konsand and Katre Õim. The typological work of the phrases that was started eight years ago at the paremiology department of the Estonian Language Institute was principally completed in 2000. All the different formulation variants of each phrase have been assembled; the most frequent and most characteristic of them are the type titles, or the headwords, where the literary form is preferred (*naar niu hamba pill suust maha* 'laughed his/her teeth out', *hambad kukkusid naermisega suhu* 'laughed so that the teeth fell in' – '(title)). The number of entries and their popularity are not really associable. The popularity of a proverbial phrase can, in fact, be inferred from the abundance of its variations: it is commonly known that phrases are known to have a large number of variations rather than entries. A figure of speech, often used even today, *lehma lellepoeg* 'cow's cousin', i.e. 'a distant relative', occurs only in three records (incl. *meie lehma lellepoeg* 'our cow's cousin, *ema lehma lellepoeg* 'mother's cow's cousin'), but in the total of 12 semantically close records: *lehma vennapoeg* 'cow's nephew', *ema lehma vennapoeg*, *ema lehma vellepoeg* 'mother's cow's nephew', *ema lehma tädipoeg* 'mother's cow's cousin', *ema venna lehma vader* 'mother's brother's cow's godfather', *isa lehma lelletütar* 'father's cow's cousin', *oina onupoeg* 'ram's cousin', *sinu isa velle vaderi kits oli minu isa ema käes aasta aega lüpsta* 'your father's brother's godfather's goat was given to my father's mother to milk', *isa venna vaderi poeg* 'father's brother's godfather's son', *ühe isa-ema venna vaderi ristipoja tütred ja pojad* 'father's-mother's brother's godfather's godson's daughters and sons', *minu vanaisa naise tütrepoeg* 'my grandfather's wife's daughter's son', *minu onupoja tädipoeg* 'my cousin's cousin'.

The main criteria for the single-typedness of proverbs, after M. Kuusi, are the similarity of idea and core, i.e. the variants of a proverb are formed of formulations that are semantically and lexically as close as possible, while the type is formed of formulations close in content and form (Hussar & Krikmann & Normann et al. 1980: 63). These criteria are largely applicable also to phrases.

In the course of the typological work on phrases we have outlined the subcategories of phrases, the total number of types, significant variations as well as vagueness in the structure, function, meaning, etc. of phrases. In order to facilitate the typologisation we first

grouped the phrases by their most prominent structural and semantic features: similes; end-rhymed and dialogical phrases; greetings, formulae expressing gratitude and reviling; rhetorical exclamatory questions; word pairs; single word metaphors; intensifying constructs. The expressions of the same figure are therefore classified under different types. It would be ideal if the occurrence of any figure could be observed through all the phrases. For example, the figure *saksa sool* occurs both as a metaphoric notion *saksa sool* 'ashes' as well as in the expression *saksa soola kukkuma* 'the food fell down'; the asyndetic word pair *suud-silmad* 'eyes-mouths' occurs in several hyperbolic expressions: *suud-silmad ammuli peas* 'eyes-mouths open', *suud-silmad pärani lahti* 'eyes-mouths wide open', *suud-silmad laiali* 'eyes-mouths wide', *suud-silmad häbi täis* 'eyes-mouths full of shame', *tehti suud-silmad häbi täis* 'eyes-mouths were filled with shame', *valetab suud-silmad täis* 'lies fill eyes-mouths', *lakkus suud-silmad täis* 'drank his eyes-mouths full' (of alcohol), *vikkis kõik suud-silmad lund täis* 'cheated all the eyes-mouths full of snow'. One and the same figure, for example, *naiste tahk*, *kana ninarätt* (both meaning "the floor") often crosses the genre limits *naiste tahk ja kana ninarätik*, *naiste tahk* – phrases; *naiste tahk ja kana ninarätt om vanapaganal teadmata*, *naiste tahk ja kana suurätt on üks*, *naise tahk ja kana suurätik on põrand* – proverbs; *Mis on naiste tahk ja kana ninarätt?* – riddle. This is also one reason why the organisation of phrases cannot be based on riddles. A solution here would be references to other short forms, i.e. to the corresponding type numbers in the publication of Estonian proverbs *Eesti vanasõnad* and that of the Estonian riddles *Eesti mõistatused*.

3.1. The limits of phrase type

Phrases are subject to extensive variations in the lexical, the syntactic as well as the semantic plane. While the total number of phrase types is close to that of proverbs and riddles, the phrases differ quite considerably in the other aspects: the number of entries in a type is smaller and the number of types with the same formulation variants is probably the largest (Krikmann 1997: 185). The borders between different phrase types are particularly vague.

The most problematic aspect of the limits is the synonymity of phrase constituents. It is difficult to delineate the border between the formulation variants of the same phrase and synonymous expressions. Feliks Vakk mentions phraseological variants of semantic, structural and lexical unity. Differences are only in the lexical and grammatical formulation of single components. F. Vakk basically distinguishes between synonymous (structural synonyms) and semantically differentiated variants (Vakk 1970: 268). The lexical components of phrases, however, are often not linguistic synonyms but synonyms based on context, which manifests both on the level of meaning as well as of form. The speaker's freedom of choice in modifying and substituting, etc. phrase constituents is quite considerable. Here lie the roots of most problems related to defining a phrase, namely, whether an expression is another variant of the same phrase or an independent, though synonymous expression.

The following phrase type demonstrates the variation of *grain* ('iva')– substantive (*iva > iva-mari > mari > tera > tang > tangutera > tilk ja tang > raasuke > leivaraasuke > teraraasuke*): *pole **ivagi** hamba alla saanud* 'have not gotten a grain under the teeth (have got nothing to eat)', *pole veel mitte **iva** suhe pistnd* 'have not put a grain in the mouth', *ei olõ viil jumala **üvvä** suuhtõ saanu* 'have not yet gotten a God's grain in the mouth'; *ei ole **iva marja** hamba alla saanud* 'have not gotten a grain's berry under the teeth', *põle veel amma alla saand ei **iva ega marja*** 'have not gotten neither grain nor berry under teeth', *miu suhu ei ole mitte **iva marja** ka saanu* 'no grain berry has gotten into my mouth', *ei olnud **iva marja** kah näinud* 'have not seen a grain berry'; *ei ole veel **üvvä ei marja** maitsanu* 'have tasted neither grain nor berry'; *ei ole veel **marjagi** hamba pääle panden* 'have not put a berry on the teeth', *ei ole veel **marja** suhu saanud* 'have not gotten a berry in mouth', *mette **marja** maitsend* 'have not tasted a berry'; *olõ-õi **terräge** hamba pääle panda* 'no grain to put under the teeth', *en süönd **terägi*** 'have not eaten a grain', *põle mitte **teragi** süüa* 'not a seed to eat', *mitte **teragi** ei jäetud* 'no seed was left', *mitte üts jumala **tera*** 'not one God's seed; *põle mette **tangogi** suho sand* have not gotten a grain in the mouth', *mitte taevalist **tangu** suhu pista* 'not a heavenly grain to put in the mouth', *mitte jumala **tangu*** 'no God's grain', *ei antant mulle mitte **tanku terägi** ampa pale* 'I wasn't given even a grain's seed under

my teeth', *pole veel **tilka tangu** sohe sand* 'have not gotten a drop of grain in the mouth'; *mitte **raasukestki*** 'not a crumb', *põle **leeva-raasukest** amma alla suand* 'havenot gotten a bread's crumb under the teeth', *võttis ära viimase kui **teraraasukese*** 'took away the last grain crumb'. Also, this is a case of verb substitution: *pole saanud > ei ole pandud > ei andnud > ei ole näinud > ei ole maitsnud > ei ole söönud > ei ole jäetud > ei ole pista > võttis ära* 'have not gotten >have not been put > have not given > have not seen > have not tasted > have not eaten > have not been left > have nothing to eat > took away', also the structure of the expression and the subject of the activity: *en süönd terägi > põle mitte teragi süüa > võttis ära viimase kui teraraasukese* 'did not eat a grain > have not a grain to eat > took away the last grain crumb'.

The following phrase type is one of the ten sayings with the largest number of variants. Semantically close expressions about taking confirmation differ both in the lexical as well as syntactic aspect, the dominant formulation here is *papi sigu söötma* 'feed the minister's pigs': *papi sead söödetud* 'the minister's pigs are fed', *kirgu sead öpetaja juures äe söötnd*'church's pigs fed at the minister's', *kerik-ärra sigu söötmas* 'church minister feeding the pigs', *käis kirikus sigu söötmas* 'went to the church to feed pigs', *õpetaja sead ää söötud* 'minister's pigs are fed', *köstre sigu söötma* 'to feed parish clerk's pigs, *köstre tsiku lugema* 'to read parish clerk's pigs', *aja inne papi tsia lauta* 'first drive the minister's pigs into the sty', *jooda köstri vasikad ära* 'water the parish clerk's calves', *papi sead alles söötmata* 'minister's pigs are not yet fed', *papi tsia pahta ajamada* 'minister's pigs are not driven to the sty', *köstre kapuste alle söömäde* parish clerk's cabbages are still not eaten', *alles köstri sead söötmata ja papi kapsad söömata* 'parish clerk's pigs are still not fed and minister's cabbages not eaten'. The second split concerns the figure *seaküna ümber lükkama* 'to push over the pig trough': *läks seaküna ümber lükkama* 'went to push over the pig trough'; *lükka enne seaküna umber* 'first push over the pig trough', *lükka enne kiriku herra seaküna umber* 'first push over the minister's pig trough'; *Kas said papi seaküna ümber?* 'Have you pushed over the pig trough?'; *Kas sul juba kirikärra sead söödetud ja küna ümmer lükatud?* 'Have you fed the minister's pigs and pushed over the trough?'; *nüüd on papi seaküna püsti* 'the minister's pig trough is now up', *õpetaja*

seaküna ümber lükand 'pushed over the minister's pig trough'; *köstri seaküna püsti aamata* 'parish clerk's pig trough is left pushed over', *seaküna alles ümber lükkamata* 'pig trough not yet pushed over', *seaküna ots on kergitamata* 'the end of the pig trough is not lifted', *seamold alles ümber lükkamata* 'pig tray still not pushed over'; *siasaba on sirutamata* 'pig's tail is not straightened'; *aja enne siasaba sirgeks* 'first straighten the pig's tail'. Similar extensive clusters often seem to require the division into smaller units, even into different types.

An altogether different problem is longer expressions containing two or more type rudiments: *Mis saab sest Murile murda või hallile anda!* 'What's there for Buddy to kill or a wolf to give!', *Mis sest saab Murile murda või hallile anda või kirju koerale enesele süüa!* 'What's there for Buddy to kill or a wolf to give or a spotted dog to eat!', *Ei sest saa Murile murda ega hallile anda*, *kui omalgi peenikene pihus* 'There's nothing for Buddy to kill or a wolf to give, if you're in a pinch yourself'. Single texts that first appear as contaminous prove to be random accumulations of traditional phrases: *omal alles kõrvatagused märjad, piimahambad suus ja nokk kollane, aga tahab teistest targem olla* 'still wet behind the ears, milk teeth in his mouth and green horns, but thinks he's smarter than others'; *keerutab nii ja naa, võlsip suu-silma täis ja lätt esi kos tuhat* 'tells things in a roundabout way, fakes your eyes-mouths full and ; *panõ koli kotti, hamba varna ja asi tahe* 'put his things together, "hung up his teeth"(i.e. had nothing to eat) and that was it'. Simultaneously, set phrases may occur as independent types (*pane vai jänessega juuskma* 'let him run with hares'*, näie ära, kui kihulane kerku tornin aegut ja kuulse tolle kah ärä, kui kirp üüse sängikoti õlgi sissä kusse* ('he could see when a midge yawned at the church tower, and could even hear when a flea pissed into the straws of the pallet at night') or as contaminations: *Kui mina nuur olli, kül mina olli virk: joosi jänessele järgi, kuulin kui kirp õlgi sisse kuss, näie ära kui kihulane kirko tornin haigut.* 'When I was young, I was so deft: ran after a hare, heard a flea pissing in the straw, saw how a midge yawned at the church tower' *Kül ma olle latsen tragi, ma joosi nii kõvaste, et jala es putu maa külge, ja kui teräne ma olle, ma näie ära kui kihulane kerku tornin aegut ja kuulse tolle kah ärä, kui kirp üüse sängi koti õlgi sissä kusse*. 'I was so nimble when I was a child, I ran so hard that my feet never touched the ground, and I was so smart that I

saw how a midge yawned at the church tower and even heard, how a flea pissed into the straws of the pallet at night.')

Since typologisation is largely negotiable and the result often subjective and the variation range of phrases relatively wide, a question rises: is it really necessary to group together expressions of the least similarity and fit them under types? (Well, unless typology is a goal in itself.) While dealing with card files the grouping of expressions into types is definitely justified, since this would be the only way to find expressions of similar structure and meaning. An electronic database also enables searches by other criteria.

3.2. Headword

A headword is the first meaningful single word granomen of an expression that functions as a title text. Due to the peculiarities of the material we have presently realised that a headword selected by this principle cannot fulfil its purpose and function as the central figure of the expression. This principle of defining a headword cannot comprise semantically and lexically similar expressions, they simply cannot be found. The following expressions of similar syntactic structure and lexical constituents, for example, have been given a different headword: *ta oli juba külm* – headword *külm*; *tema kannad on siis külmad* – headword *kand*; *tal omma juba käe külmä* – headword *käsi*; *sel joba külmakenga jalan* – headword *külmakingad*. The type *aja kui väsinud hobust hange* is located under the headword *hobune*, the type *nagu veohobune* under headword *veohobune*, the type *nagu hobuse selg* – headword *hobuse selg*, the type *otsib hobuse seljast* – headword *hobuse seljast*. This poses a question of the efficiency of such headwords. Considering the structure of the future electronic database, it hardly seems to have any need for a lexical headword. The existing semantic slots could, in principle, be kept together by the entitling of expressions with the figure, i.e. by decomposing it, e.g. the expression *kannad on külmad* 'heels are cold' is decomposed to *kannad külmad* 'heels cold'.

Clearly, the observation of the structure and meaning, single word or word combinations of the phrases in this corpus (the phrase files) is only possible by manual browsing through the whole material. Presently, the search by the phrase structure only gives the follow-

ing similes (*helde nagu Helde-Mari poeg* 'generous as the son of generous Mary'; *justkui koer heinakuhja otsas: ei söö ise, ei lase teist süüa* 'like a dog on haystack: does not eat himself and does not let the others either; *kaalub kui kulda* 'weighs as much as gold'; *ole kui küpsiküüsilaste juures*; *seab ja sordib nagu juut loeb oma kasukopikaid* 'sets and sorts like a Jew counting his coins') and intensifying constructs (*kõht on nii tühi et sööks kas või pool hunti korraga* 'I'm so hungry, I could eat half a wolf'; *läheb nii et kivi kuuleb, teine näeb* 'treads so that a stone can hear and another one can see'; *valetab nii et suu suitseb* 'lies so that his mouth smokes') that, as was previously mentioned, were set apart before the typological work for purely technical reasons. Less productive structural phenomena can be observed only when the phrases with necessary features are searched individually in the files. Nominative phrases like *ike seljas* 'yoke on the shoulders', *kananahk seljas* 'gooseflesh on the back', *karvad turris* 'hairs [standing] up', *keel sõlmes* 'tongue tied', *keel suus* 'tongue in the mouth', *kilk kaelas* 'cricket around the neck', *kirp kõrvas* 'flea in the ear', *kirves kotis* 'axe in the bag', *kits kotis* 'goat in the bag', *kitsed orasel* 'goats in the plant shoots', *kolk kaelas* 'swingletree around the neck', *koorem küljes* 'a burden attached', *krapp kaelas* 'cowbell around the neck', *kuri karjas* 'fat is in the fire', *kõlkad kõhus* 'chaff in the stomach', *hammas verel* 'bleeding teeth', *lambanahk seljas* 'sheepskin on', *maik suus* 'taste in the mouth' are related in that they all indicate to the object's condition or one of its characteristics (*kits kotis* "out of money", *nälg näpus* "in low water", *kilk peas* "drunk"). One of the phrase components, substantive in nominative, functions as a subject, the other as an adverbial; the adverbial component often occurs in locatives (Baran 1999: 59). Since nominative phrases are positioned under different headwords far apart in the files, the study of (phrases of) certain structure in the whole material is relatively problematic and time-consuming. Here are some examples of conditional phrases: *hakkab ujuma, kui vesi perse puutub* 'he will swim, when his ass gets wet'; *kui hurdast saab karjakoer, siis saab temast ka inimene* 'if a greyhound becomes a herding dog, it will also become human'; *kui jumal pole lätlane, läheb see korda* 'if God is not a Latvian, all will work out'; *küll siis nägijaist saavad, kui silmad pähe tulevad* 'they will see, when eyes will grow on them'; *küll siis teeb, kui omad kirbud sööma hakkavad* 'he will do it, when his own fleas start eat-

ing him'; *kui sa mu uksest välja ajad, tulen aknast tagasi* 'drive me out the door, I will come back through the window'; *kui sai tarre, siis tahtis juba tareharja* 'was let into the house, already wanted the roof' (for other examples see, Baran 1999: 64). Obviously, we don't need to add that in conditional sentences like this the designation of a headword is extremely problematic, if not impossible. Not to mention the fact that often the somewhat randomly chosen content-lexical headword fails in grouping phrases of similar structure.

The frequency study of single word or word combination is possible if a certain word or word combination acts as the core of expression, the central figurative element, figure of speech or a set of notions (see Hussar & Krikmann & Normann et al. 1980: 63), and also as the headword. In order to be able to observe words that are less significant in the expression as a whole, the material has to be designated syntactically, semantically or in some other way.

4. THE DATABASE OF THE ESTONIAN PHRASES

According to ISO a database is a collection of data organised according to a conceptual structure describing the characteristics of these data and the relationships among their corresponding entities, supporting application area(s) (Estonian Standard EVS-ISO 2382). The database of the Estonian phrases is a composite of different categories within the material which enables to acquire various kinds of information on the phrases; one of the aims of establishing the database was also making preparations for the draft of the printed publication of the Estonian phrases.

In a sense the database was being constructed already during the typological work of phrases – the type titles and common formulations were entered into the initial database. The latter formed the basis for the electronic search engine at http://haldjas.folklore.ee/rl/date/robotid/leht3.html (programmed by Sander Vesik, the Estonian Literary Museum), which provides information for nearly 25.400 Estonian phrases. The material is copied from the dictionary of phraseology *Fraseoloogiasõnaraamat* (1993) by Asta Õim (nearly 7.300 expressions) and the manuscript materials of the Estonian

Folklore Archives and the dialectal archives of the Estonian Language Institute (nearly 18.100 expressions altogether, incl. approx. 10.600 similes). The search can be conducted by a part of the word, the string **kass** may give results *kassi, kassa* as well as *okassiga*, but also by the matching word: **(kass)**, the search engine also enables the use of alternative letters in search: **nai[ns]e**. To search for combination of different words requires the use of markers of logical expressions: & 'and', | 'or', i.e. the search for **nai[ns]e & kass** finds all the expressions that contain the words *kass* ja *naine (naise)*; the search **nai[ns]e | kass** gives all the expressions that contain the words *kass* or *naine (naise)*. In order to find phrases that would contain words like *kass* or *koer* and *nagu*, the string that you need to insert is: **{kass | koer} & nagu**. The search engine relies on formulations and carries no information on archival data; similarly to the paper files the structure, meanings, etc. of phrases cannot be studied before singling out the expressions one by one.

4.1. Database structure

In order to decide which information must be included in the database, we first have to choose the criteria and characteristics by which the expressions can be found, treated and analysed. The most easily accessible information should be the data on the phrase distribution and productivity, as these data have already been included on the index cards or can be directly derived from them. In order to communicate information on the form, structure, figurative semantics, meaning, etc. each phrase type must be studied from a certain aspect and marked accordingly. The presentation of the material in the database has to be good enough to be successfully used, say, by a scholar interested in expressions of certain structure, as well as anyone, who wishes to find expressions of the established meaning. In other words, the search engine must find all the expressions that include, say, a substantive, verb and illocative: *kali läheb kaevu* 'kvass goes to the well' – *, leib hakkab jalgu* 'bread goes to the feet', *mõte kargas pähe* 'an idea came to the head', *mokk ulatub klaasipõhja* 'mouth reaches all the way to the bottom of glass' – 'a heavy drinker', *nälg tuleb näppu* 'hunger will come to the fingers', *näpud puutuvad põhja* 'fingers touch the bottom' – 'to be broke', *pada läheb pajusse* 'pot will go to the willows', *päevad lähevad sulase poole* 'days

will go to the farmhand' – 'days are drawing to an end', *silm läheb looja* 'the eyes will set' – 'doze off', *silmad jäid õue* 'eyes were left outside' i.e. coming from the bright light the eyes take time to adjust to the darker room, *silmad löövad sirklisse, suu jäi lukku* 'mouth was locked up' – 'became silent', *suu läks hukka* 'mouth went bad' (see Baran 1999: 50), or those expressing ignorance: *targad kui Saaremaa varesed* 'smart as the crows in Saaremaa', *tark kui sandi kark* 'wise as a lame man's crotch', *pea nagu sõel* 'head like a sieve', *ei tea midagi nagu siga pühapäevast* 'knows nothing like a pig knows about Sunday', *nagu siga ajab molli umber* 'pushes over the trough like a pig', *nagu siga joob igast porilombist* 'drinks from every puddle like a pig', *nagu siga sööb soolaga ube* 'eats salted beans like a pig', *nagu siga taevatänavasuus* 'like a pig on the heavenly street', *nagu siga tuhnib põlluhunniku asemel* 'grubs in the field like a pig'.

A good example here would be a formulation *jalad justkui hane lestad* 'feet like goose webs' that belongs under the simile type *nagu hane jalad* 'like goose feet. Data may be tentatively divided in two: objective data and subjective data. Objective data comprise all the information included on the index cards (unfortunately often incomplete):

- **archive reference:** RKM II 340, 305 (478),
- **place of collection:** Emmaste parish
- **collector:** J. Kõmmus
- **informant:** missing (like in most of the records)
- **year of collection:** 1979
- **"**the main part", i.e. **the phrase formulation** independently, usually also in context

comments, meaning, etc: *hanelest*: *Jälad just'kut ane-lestad (Öeldakse lapse kohta kelle jalad on külmetamisest punased)*. Goose web: Feet like goose webs (indicating to a child's feet, which are red of cold.)

Subjective data is based on information not directly included on the index card, but which have been learned or established on the structure, lexica, semantics, etc. of phrases. Typological and descriptive data belong here.

-    **Headword**: *hani* ('goose')

We have already learned that due to the peculiarity of the material as well as the relatively randomly selected criteria of selection the headword as such cannot be efficient and function as the central figure of the expression. Headword also fails to bind semantically and lexically similar expressions, the search of which within the whole corpus is presently extremely time-consuming. Different semantic slots could be grouped more closely by designating the expressions by the figure, i.e. decomposing the figurative headword (e.g. *kannad külmad* 'heels cold' in the expression *kannad on külmad* 'heels are cold').

- **Title text of the phrase type:** *nagu hane jalad* 'like goose feet', i.e. the most characteristic, frequently occurring formulation among the members of the phrase type.

- **The original formulation:** *jälad just'kut ane-lestad* 'feet like goose webs'. One might wonder why formulation has been classified as typological data if it's already included among the objective data. There are many reasons for it: firstly, isolating the formulation from the context makes the analysis technically simpler. Secondly and more importantly, the limit of a phrase may be very vague and diffusive – unless the formulation appears independently (i.e. without the context) on the card, then the beginning and end of the phrase may not be fully recognised. For example, whether the formulation of the sentence *Pueb nindägu porsas pohust väljä, touseb magamast üles ja jädäb vuodi kohendamada* 'Crawls out of bedding straw like a pig, wakes up and leaves the bed unmade' is *nindägu porsas pohus* 'like a pig in the bedding straw' or *pueb nindägu porsas pohust väljä* 'crawls out from the bedding straw like a pig', i.e. is the verb *välja pugema* 'crawl out' a constituent of the phrase or belongs to the context? Since the defining of formulations is often a matter of negotiation, similarly to the headwords and type titles described above, then the user of the database should be aware of their (partial) subjectivity.

- **Cross-references between phrase types and/or formulations:** see *punased kui pardi jalad* 'red like duck's feet', see *punased kui kure jalad* 'red like stork's feet'.

-   **The sequence of a given formulation within the phrase type** to describe the interrelations of the different formulations of the same phrase type on the basis of their lexical, syntactic and/or other characteristics (see also EV 1980: 68, 69).

Descriptive data comprises the characteristics of the phrase in its entirety, as well as those of its constituents – the following list, however, is hardly finite. Descriptive data should also include information on the figure of a phrase.

-   **Description of formulation structure**

-   **The meaning, explanations and comments of formulation:** *lapse kohta, kelle jalad on külmetamisest punased* (About a child whose feet are red of cold).

-   **The different semantic description of formulation constituents:** *hani* 'goose'*:* + animate, - man; *lestad* 'webs'*:* - animate, etc.

4.2. Application of the database of phrases

The establishing of the concise database of phrases was initiated in 1998. Presently, the material has been entered into the computer, also the objective and most of the typological data (incl. originally formulation, title text and headword of a given phrase type, references to analogous formulations or types) has been added without major revisions. Additional descriptive data includes the meaning of formulations, in case it is presented or explained in the index card text. Similes are supplemented with the description of the title text structure of each type and the semantic description of title text constituents, also the meaning of less known (dialectal) words in formulations; on some occasions the sequence of formulations in simile types has been established.

The electronic database of the Estonian phrases is financed by the Estonian Science Foundation and the Estonian Culture Foundation. The search engine, which meets the present needs for analysis and aims for the nearest future, has been completed (programmed by Indrek Kiissel – Estonian Language Institute). The database provides a user interface, enabling to edit archival data,

check the authenticity of phrases, approximate typology, define the structure of phrases, etc. The basic functions of the database are the following:

- **Communication of data.** The database communicates data depending on the search type (multiple searches and logical expressions), enabling searches based on the data as well as the collector. The search can be conducted through one or several fields and results arranged in a certain order.

- **Adding, editing, deleting of data in real time.** The entering of data for operating or saving, editing. This function enables to add and remove entries, edit archival data and, if necessary, upgrade/add, revise and synchronise title texts/formulations of types, synchronise the marking of dialectal peculiarities, detect spelling errors, etc. The use of the database in real time enables several users to simultaneously edit data and conduct searches.

4.3. The quantity or data and the quality of search results

We have already emphasised that the idea of the database is not to preserve data, but it is intended to present as much additional information as possible, i.e. it should include directly or indirectly the different results of syntactic, semantic, etc. analysis.

The access to complete and objective information in the database is largely obstructed by the use of dialect and the quality of formulation, which no longer meet the rules of contemporary Estonian orthography. Obviously, the search for expressions containing, say, the word *harakas* 'magpie', results in exact matches of the phonemic sequence: *edev nagu **harakas**, **harakas**eks minemä, kalts-**harakas**, kädsatab nagu **harakas**, laseb saksa keelt kui **harakas**, nagu kõrbend **harakas**, paju**harakas***. The rest of the entries containing a declension, dialectal form, misspelling, etc. of the same word, such as *kerge kui **harak**, kirju jusgu nougi**harak**, nagu **haraka** pesä*; *armastab nagu kiits**arakas** läikivaid asju, kisendab justkui **arakas** aiateibas, kädistab ku **arakas**, lendab kui **arakas**, nigu **arakas** üppab, naerdasse nagu **arakast** ajateibasse*; *kiitsagas nidagu **arak** aiateibäs, kädsäts ku tuli**arak**, nagu **arakud** aiateibas,*

*paio**arak**, pallas nigu **arak,** pelglik ku **arak**, vahib **arakaviisi** otsa* cannot be found. In other words, the search is hardly accurate.

Which descriptive data should then be needed for a fully functional database? Does the search word have to be perfectly accurate (i.e. correspond to the result), or should the dialectal formulations (i.e. all the phrase constituents) be provided with "translations" into the literary language? Below we will present some solutions.

If the original formulation in each entry is provided a correspond-ing literary form (*senusuguine sagataja pane vai arakaga paari – sinusuguine sagataja pane või harakaga paari*, *kuiv nagu kiitsaka tiiv – kuiv nagu haraka tiib*), then the work would become ex-tremely laborious. One of the problems encountered is the lack of corresponding literary words: *nagu Kristuse piibu**kriiska**. Also, making such amendments requires much routine work; the ex-amples above indicate that very many amendments must be made on similar words. Moreover, (dialectal) words that are unfamiliar to the modern user require translations. This means defining the words and presenting the meaning (it has been done with some similes); as to the polysemic words a reference should be made to the more accurate meaning, cf. *sagattab nindagu aragas aia teibäss*: sagatama **'kädistama'** (cackle) (Pall 1989: 377), 'jutuga vahele tulema, vahele segama'(intefere), *vareste vaakumine*, *harakate sagatamine* (Saareste 1958: col. 220, 610); *must nagu aanikas*: aanikas 'kännutüügas; langetamisel tüve kännupoolne ots'(tree stump), 'langetamisel tüvesse raiutud sälk'(notch hacked into the treetrunk at felling), **'leotus-, keedu- või pesuvesi'**, 'õllevirre, õlu' (Juhkam & Must & Mäger 1994: 53–54) (light beer, beer), 'tüügas' (Pall 1989: 12) (stump). The easiest solution would be ref-erences to the electronic publications, like the index of the dialec-tal dictionary *Väike murdesõnastik* and the semantic dictionary of the Estonian language *Eesti keele mõisteline sõnaraamat* by A. Saa-reste. The idea of presenting parallels in the literary language is to give objective results to the search in contemporary Estonian. At the same time, is this intermediate level that calls for copious extra work really necessary, especially if some of the words do not seem to require parallels in literary language:*jusku emmis*: sööb *isi* oma põrsad ära*, käib järel nagu **eilane** päev*, rumal kui **iesel**, röögib **nigu eläjäs** ('like a sow: devours her own piglets, follows

like "yesterday", dumb as a donkey, growls like a beast') And also remains the problem of formulations, i.e. the search of expressions including the word *jalg* 'foot', 'leg' results in both

> jalg (SgN):***jalg*** *nindägu kassi kaul* 'foot as a cat's neck', *kena kui keripuu****jalg*** 'lovely as a swift foot', *suur **jalg** nagu hobusel* 'big foot like a horse's', *nagu ark****ialg*** 'foot like astride'

as well as

> jalga (SgP): *läks jänese****jalg****a* 'went like with a hare's foot';
>
> jalge (PlG): *pitk ja kitsas noagu **jalg**enarts* 'lean and lank like a foot cloth';
>
> jalgu (PlP): *kõht täis nagu Nõva mehel jõululaupäevaõhtu orava päid-**jalg**u* 'stomach is full like a Nõva man's stomach of squirrel heads and legs on a Christmas Eve';
>
> jalgus (PlIn): *nagu kana **jalg**us takkud ei soaa emmale ega kummale* 'like a hen with tow in its feet cannot have either'.

In order to find a word in any possible form, the expressions should be divided into grammatical units and entered into an index of original words: nouns with the singular nominative root, verbs with the *ma*-infinitive. (Otherwise the search would have to cover the whole paradigm and even then might not yield objective results). From the index the search would take the user to the original formulations. A search by the word *jalg* ('foot', 'leg') would give the results above and the following:

> jalaga (SgKom): *laisk nigu kolme **jalaga** vana obene* 'lazy as an old horse with three legs';
>
> jalad (PlN): ***jalad** kanged nindagu ange aetud obusel* 'legs stiff like a horse's in a snowdrift', ***jalad** nagu jahumatid* 'legs like mats of flour', *jusku **jalad** all* 'like has feet', *nao va ane **jalad*** 'like goose feet', *sesab **jalad** laijali nagu ark* 'stands legs astride like a fork', *vereva kui hani **jala*** 'red as goose feet'.

The main advantage of search by the word index would be the possibility to disregard spelling errors and the morphological form of a word.

Important information on grammar could be added to the database by adding a morphological index to the words of the word index. Of the possible components of a morphological index used in *A Concise Morphological Dictionary of Estonian* (1992) by Ülle Viks it would be practical to use the marker of the word class and type numbers for variable words, suggesting the need for type description in the chapter dealing with grammar. Generally speaking, the phrases could be handled as follows: *sagattab nindagu aragas aia teibäss* > *sagatama* V(erb) type no. 27, *harakas* S(ubstantive) 2, *aed* S 22, *teivas* S 7.

In order to find a specific word form, the words in expression need an additional individual morphological index describing the given form of transformation: *lapsi kui järijalgu* '[has many] children like seat legs' > laps 'child' S 14 PlP, järi 'seat' S 17 SgG, *jalg* 'foot', 'leg' S 22 PlP. In this case the search should include the word in its original form (*laps*) and the morphological description of the searched form: PlP.

I. Kiissel, the informatics administrator of the Estonian Language Institute conducted an experiment that included nearly 40.000 formulations. A morphological analyser compared the words automatically with the material of the dictionary of form. The result was the total of 39.035 different language variants consisting on the average of 7 letters. Owing to the dialectal peculiarities and spelling mistakes mentioned above, the number of variants included in the analysed phrases is considerable. Cf. 'horse'

   hobune (SgN) – hobene, hobõnõ, hobu, hobbo, hobus, hobes, hobos, hobbus, hobbos;
   hobuse (SgG) – hobbuse, hobose, hobbose, hobese, hobõsõ;
   hobust (SgP) – hobõst;
   hobusele (SgAll) – hobõsõlõ;
   hobusel (SgAd) – hobõsõl;
   hobusega (SgKom) – hobõsõga;
   hobuste (PlG) – hobbuste, hobeste;
   hobuseid (PlP) – hobesit, hobusi;
   hobustega (PlKom) – hobõstõga.

In the contemporary Estonian language the word is known to have 33 variation forms (Viks 1992: 7). This indicates that the number of original forms in the future word index as well as the lexica of the material of the Estonian Folklore Archives is limited. The same is confirmed by the analysis of a hundred randomly selected variants, which is reduced down to 40 words in original form. The original forms should be supplemented with a morphological index (word class marker + type number).

a I 41: *juut oll kauplema a härrä oll masma*; *oledki sina see a ja o*

aaberjuut S 22: *aaberjuut siakõru*, *siakõru*

Aabraham A 2: *aabrahamme nännö*; *Abrahaame nännü*; *aabrahammiga kuoh*

Aabram A 2: *Aabrami sülem*; *Aabramid jälle näind*; *Abramid näind*; *Aabramist nännu*; *aabramet nännu*; *Aabramit nägema*; *näitas teisel Abramit*; *aabrammi näind*; *Abrami näind*; *aabramiga ka kokku juhtund*

Aadam A 2: *Aadam ja Eeva läksid üle Neeva*; *Aadama ajal*; *aadami ülikond*; *aatami kahvlid*; *aatama orkid*; *aadamast ja iidamast*

aadamaaegne A 2: *Aadamaaegne*; *aadama-aeksed*

aadamakahvel S 22 ~ S 2: *Aadamakahvel*; *aadamakahvlega*

aader S 2: *aadred laskma*; *aadrid laskmas*

aamen I 41 ~ S 2: *aamen konks ja hapud silgud*

aanispits S 22: *aanispitsid teha*

aarabard S 22: *Ats Aarabard ja Juks Jõuluvorst*

aarna S 1 ~ V 29: *astume Aarnale, müttame Möksi, tuleme Tillele*

aas S 22: *pole veel aasa peele saand*

aaspuk S 2: *aaspuk*; *aaspuuk*

aasta S 1: *aasta selgas juba*; *aastad kipuvad õhtule*

aastak S 2: *aastak ja ase, kuu ja kortin*

abakukk S 22: *abakuk*

abi S 17: *abi oli kohe marjast*; *abi saa kennei arva, tuge saa kennei tunne*

abi-armu S 22 ~ V 27: *abi-armu pole kohegilt tulemas üht*

abukasukas A 2 ~ S 16: *abukasukas*

abus S 17: *aastud abus*

ader S 24: *adra alla teind*

adramadrus S 9: *adramadrus*

adramees S 15 ~ S 13: *adramees*; *adramies*; *adramiis*; *adramiis*

adru S 16: *adru ainiti*

ae I 41: *Hans, ae aja härjad koju, homme kroonupüha*

aed S 22: *pannasse sabapidi aateivasse*; *meie räägime aeda sina räägid aea auku*; *kärner aedas*; *tapu aeast läbi tulnud*

aeg S 22: *aeg oo ümber*; *elud aa tuule vallal*; *Onts sul natuke aega? – Jah on küll. Mis siis? – Pane oma püksiauk kinni*; *ei ole aegu tseapahan*; *aast ja arust*

aegne A 2: *enneaegne*

ahh S 22: *ach ty psia krew Martyna Lutera*

ais S 22 ~ 23 S: *alt aesa kaeru andma*

ajaja S 1: *suksutajid pallu, aga päitse pähä aajit mitte ütegi*

ajama V 27 ~ S 2: *ahka aama*; *ahka aema*; *a'ada hambasse*; *aan talle Käojaani*; *aad vaska lassi*; *mis sa aat tühjä kotte püstü*; *aesd oma jönni*; *aest jähimihe jüttu*; *ae jälad köhu-alt välja*; *tühja kotti aead püsti*; *ära aea oma ambid irevile*; *töö aa sene kee lahti*; *aab eesa nena püsti*; *aap nigu ahka*; *kops aeab üle maksa*; *aeap peele*; *aeb kangut*; *aas Tedre Antsu*; *aase tesele kärpsit silmä*; *isi ta kusi ja kudus sukka, üleaea aeas naabriga juttu*; *aes päkad püsd*; *aame küläle kõrra päale*; *aage käe kõtust vällä*; *aavad nina kiiva*; *aasid vanamehega isi jalad segi*; *aeasid mo ka kargama*; *aedi aja ja ange vahele*

haamer S 2: *aambri alla panema*

habe S 4: *abe itsis ees*; *abä issa puälä*; *annab vasta abet*; *abemesse pomisema*; *abenasse rääkima*; *habemega kärtsi pärandat pühkima*; *abenega jütt*

haigus S 22 ~ S 9 ~ S 11: *aegos oo ea korra maha võtnd*

hea A 26: *küll sul on ia nena*

häda S 17: *adä ristmin*

olema V 36 ~ S 2: *mihel oo kroonupüha*

äär S 13: *ei sel sanal ollu aara ega otsa*; *kes tiedab kus sen aarad ja otsad jo on*.

We will hereby point out a few problems that may be encountered during the compilation of the word index. If we begin determining the original forms from the alphabetically arranged variants, the word initial 'h' or the lack of it in dialects complicates the association of different forms of the same word. The word *haamer*, for example, is sometimes written *aamer*, and also *haamer, haamert*. Determining the original form of the words of ambiguous meaning one must consider the context and consult the corresponding entries in the database. E.g. are all the linguistic variants *aa*, *aab*, *aad*, *a'ada*, *aage*, *aama*, *aame*, *aan*, *aap*, *aas*, *aase*, *aasid*, *aat*, *aavad*, *ae*, *aea*, *aeab*, *aead*, *aeap*, *aeas*, *aeasid*, *aeb*, *aedi*, *aema*, *aes*, *aesd*, *aest* traceable to the original form *ajama*? Is it necessary to interpret the obviously unfamiliar (dialect) words for a modern language user, is it the responsibility of the database at all? How far should we trace the presence of words in dialectal sources? Should the word index refer to the list of formulations containing the given word or the list of type titles? These questions have yet remained unanswered. As the word index is presumably in the literary language, the revision of the morphological index is basically the singling out of the most accurate or suitable of the issued descriptions. The morphological analyser describes the word *haigus*, or 'illness' for instance, as follows: 1) haigu SgIn > haik S 22; 2) haigus SgN > haigus S 9; 3) haigus SgN > haigus S 11.

Regardless of the potential finite scope of the original forms, the compilation of the word index will undoubtedly be complicated and time-consuming, particularly so because most of the work has to be done manually. Then again, if we use a simple structure description like word class markers in certain positions for all type heads or formulations, such as S + Adv + V: *susse püsti lööma* 'kick up the slippers' – 'to die', *silmi kinni maksma* 'pay the eyes shut', *suud kinni panema* 'shut the mouth', *suud lahti kiskuma* 'pull the mouth open', *südant rahule jätma* 'leave the heart be', *nahka lõhki lööma* 'crack the skin open', it will give us rather flexible alternatives for

studying the structure of expressions (see also: Baran 1999: 49 ff). This would enable to register all the expressions of similar structure, resulting in an exhaustive survey of structures represented in the Estonian Folklore Archives or the database, their productivity or uniqueness, etc. It would also enable to follow the frequency of the established structures, like S + V + Adv in the given material.

As a result of the previously mentioned experiment, a little more than a fourth of the 39.035 linguistic variants could be analysed on the basis of the dictionary of form with the morphological analyser (+ in Table); more than half of the variants did not have a counterpart in the dictionary, but the speculated morphological description was issued (?); about a tenth of the words could not be analysed (-); an asterisk (*) marks incorrect descriptions, correct descriptions (with certain reservations) appear in bold. We will hereby give an example analysis of fifty variants (the morphological description includes the description of the variable form, supposed original form, word class marker and type number, divider of compound words).

Variant	Corresp. word in the dict. of form	Morphological description
A	–	
aa	+	*ID > aa I 41
aab	?	*SgN > aab S 22
aaberjuut	?	**SgN > aaberjuut S 22**
Aabrahamme	?	*PlP > aabrahamm S 22
Aabrahammiga	?	**SgKom > aabrahamm S 22**
Aabramet	–	
Aabrami	?	*SgN > aabrami S 1; *SgG > aabrami S 1; *SgG > aabram A 2
Aabramid	?	*PlN > aabrami S 1; *PlN > aabram A 2
Aabramiga	?	*SgKom > aabrami S 1; *SgKom > aabram A 2
Aabramist	?	*SgEl > aabrami S 1; *SgEl > aabram A 2; *SgP > aabramis S 11; *SgP > aabramine S 12; *SgN > aabramist S 22
Aabramit	?	*SgP > aabrami S 1; *SgP > aabrami A 2

Variant	Corresp. word in the dict. of form	Morphological description
Aabrammi	?	*SgN > aabrammi S 1; *SgG > aabrammi S 1; *SgG > aabramm S 22; **SgP > aabramm S 22**; *SgAdt > aabrammi S 22; *IndPrPs > aabrammima V 28; *ImpPrSg2 > aabrammima V 28
aad	?	*SgN > aad S 22
A'ada	?	*SgN > aada S 16; *SgG > aada S 16; *IndPrPs > aadama V 29; *ImpPrSg2 > aadama V 29
Aadam	?	**SgN > aadam A 2**
Aadama	?	*Sup > aadama V 29
aadamaaegne	?	**SgN > aadama+aegne A 2**
aadama-aeksed	?	*PlN > aadamaaekne A 2; **PlN > aadama+aekne A 2**; *PlN > aadama+aekse S 6; *SgN > aadama+aeksed S 2
aadamakahvel	?	*PlAd > aadama+kahv S 22; **SgN > aadama+kahvel S 2**
aadamakahvlega	?	*SgKom > aadama+kahvle S 6
aadamakahvliga		**SgKom > aadama+kahvel S 2**
Aadamast	?	*SupEl > aadama V 29; *SgP > aadamane A 12
Aadami	?	*SgN > aadami S 1; *SgG > aadami S 1; **SgG > aadam A 2**
aadred	?	*PlN > aadre S 6; *IndPrSg2 > aadrema V 27; *SgN > aadred S 2
aadrid	+	*PlN > aader S 2
aage	?	*SgN > aage S 6; *PlP > aag S 22; IndPrPs > aagema V 27; *ImpPrSg2 > aagema V 27
aajit	?	*SgP > aaji S 16; *SgN > aajit S 2
aama	?	*SgN > aama S 16; *SgG > aama S 16; *IndPrPs > aamama V 29; *ImpPrSg2 > aamama V 29

Variant	Corresp. word in the dict. of form	Morphological description
aambri	?	*SgN > aambri S 16; **SgG > aamber S 2**; *SgG > aambri S 16; *IndPrPs > aambrima V 28; *ImpPrSg2 > aambrima V 28
aame	+	*PlP > aam S 22
aamen	+	**ID > aamen I 41; SgN > aamen S 2**
aan	?	*SgN > aan S 22
aanispitsid	?	**PlN > aanispits S 22**; *IndPrSg2 > aanispitsima V 28
aap	?	*SgN > aap S 22
aara	+	*SgN > aara S 16; *SgG > aara S 16
aarabard	?	**SgN > aarabard S 22**
aarad	+	*PlN > aara S 16
aarnale	?	**SgAll > aarna S 1**; *IndPrPs > aarnalema V 27; *ImpPrSg2 > aarnalema V 27
aas	+	*SgN > aas S 22
aasa	+	*SgG > aas S 22; *SgP > aas S 22; *SgAdt > aas S 22; *IndPrPs > aasama V 29; *ImpPrSg2 > aasama V 29
aase	+	*PlP > aas S 22
aasid	+	*PlN > aas S 22; *IndPrSg2 > aasima V 28
aaspuk	?	**SgN > aaspuk S 2**
aaspuuk	?	**SgN > aaspuk S 22**
aast	?	*SgP > aane A 10; *SgN > aast S 22
aasta	+	**SgN > aasta S 1; SgG > aasta S 1**
aastad	+	**PlN > aasta S 1**
aastak	+	**SgN > aastak S 2**
aast-arust	+	**SgEl > aast-aru S 17**; *PlEl > aast-arg A 22; *SgP > aast-arune A 10

THE MORPHOLOGICAL ANALYSIS OF LINGUISTIC VARIANTS

The reasons for such poor results are not merely the quality of the collection materials, but also the abundance of archaisms in the material. Since most of the variants are presented with more than one morphological description, it is necessary to single out the right one(s), which, again, has to be done manually.

The index of forms is efficient only if it is bound to the word index and only with the original forms of the words. For instance: **aadamakahvel S 9 SgN SgKom**, i.e. *aadamakahvel* 'Adam's fork'-'hand' is a noun belonging to the 9th variable type, appears in the database of phrases in singular nominative: *aadamakahvel* and singular comitative case (in Estonian): *aadamakahvlega, aadama-kahvliga*; the word does not appear in any other form.

In addition to finding certain word forms in the database, the form index would also prove useful at observing different grammatical categories. Cf. the expressions where the main word of the simile is substantive in comitative, performing the syntactic function of the means or subordinate adverbial: *haarab kui hammastega* 'grabs like with teeth', *käänab kui pulgaga* 'bends like with a stick', *laseb kui rabapüssiga* 'shoots like with a shotgun' – 'farts', *lõikab nagu kreissaega* 'cuts like with a disk saw', *lõikab nagu noaga* 'cuts like with a knife' i.e. 'very sharp', *otsi nagu tulega* 'look for like with a fire' i.e. 'to look very hard for something', *raiub nagu kirvega* 'hacks like with an axe' i.e.'carelessly, roughly', *torkab nagu nõelaga* 'pricks like with a needle', i.e. 'suddenly, very quickly', *võtab kui kahvaga* 'takes like with a scoopnet' i.e. 'takes a lot'; *laseb nagu ratsahobusega* 'rides like with a riding horse', *laseb kui veoga* 'rides like with a long pulling', *läheb nagu viie paari härgiga* 'goes like with five pairs of oxen', *läheb nagu lepase reega* 'goes like with an alder sleigh' i.e. 'very smoothly', *läheb nagu masinaga* 'goes like with a machine', *läheb kui rohega – nagu emis põrsastega* 'goes like a sow with her piglets', *nagu Kiki Jüri oma kanaga* 'like Kiki Jüri with his hen', *justkui Kärsa Krõõt oma tegevusega* 'like Kärsa Krõõt with her work', *tuleb välja kui Tuhalaane mees leivaga* 'comes

out like a man from Tuhalaane with his bread' (see also Õim 1997: 49).

The organisation and systematisation of the Estonian phrase files was started manually, the same basically applies to the typological work. Presently, nearly all the material of the files has been entered into the database. Certainly, the typology based on the files requires additional revision and synchronisation, also archival data has to be edited; in the nearest future we will begin segregating non-authentic expressions. The most laborious task is definitely ascertaining the meaning(s) of expressions. Hopefully the established indices will help to solve many problems. But most importantly, the practical significance of the database of phrases is that in the years to come everyone interested is guaranteed an easy access to the bank of Estonian phrases.

Comments

[1] All example texts here and below have either literal or literal and semantic translations.

References

Baran, Anneli 1999. *Eesti fraseoloogia põhijoontest ja struktuurist* [On the Principles and Structure of Estonian Phraseology]. MA thesis. Tartu: The Estonian and Comparative Folklore Department of the University of Tartu.

Eisen, Matthias Johann 1913. *Meie vanahõbe. Sarjatäis Eesti rahva endist tarkust, kõnekäänusid ja ütlusi* [Our Old Silver. A Series of Estonian Folk Wisdom, Phrases and Sayings]. Tartu: Hermann.

Erelt, Mati & Kasik, Reet & Metslang, Helle & Rajandi, Henno & Ross, Kristiina & Saari, Henn &. Tael, Kaja & Vare, Silvi 1993. *Eesti keele grammatika* II: *Süntaks* [The Grammar of the Esotnian Language II: Syntax]. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.

Estonian Standard EVS-ISO 2382 = *Eesti standard EVS-ISO / IEC 2382. Infotehnoloogia. Sõnastik* [Estonian Standard EVS-ISO 2382. Information technology. Dictionary] *http://ee.www.ee / ITterminid /* (20 December 2003).

Helle, Anton Thor 1732. *Kurtzgefasste Anweisung zur Ehstnischen Sprache*. Tallinn: Eestimaa Konsistooriumi kirjastuskassa.

Hupel, August Wilhelm 1780. *Ehstnische Sprachlehre für beide Hauptdialekte*, *den revalschen und dörptschen, nebst einem vollständigen Wörterbuch*. Riga & Leipzig: Hartknoch.

Hussar, Anne & Krikmann, Arvo & Normann, Erna & Pino, Veera & Sarv, Ingrid & Saukas, Rein 1980. *Eesti vanasõnad* 1. 1–5000 [vanasõna] = *Proverbia Estonica*. Monumenta Estoniae Antiquae, III. Tallinn: Eesti Raamat.

Hussar, Anne & Krikmann, Arvo & Saukas, Rein & Voolaid, Piret 2001. *Eesti mõistatused* I. 1–1350 [mõistatust] = *Aenigmata Estonica* I. 1–1350. Monumenta Estoniae antique, IV. Tartu: Eesti Keele Sihtasutus.

Juhkam, Evi & Must, Mari & Mäger, Mart & Neetar, Helmi & Nigol, Salme & Niit, Ellen & Oja, Vilja & Pall, Valdek & Ross, Eevi & Univere, Aili & Viires, Helmi (koost.) 1994. *Eesti murrete sõnaraamat* I: 1 (A–J). [The Dictionary of Estonian Dialects I: 1] Tallinn: Eesti Keele Instituut.

Krikmann, Arvo 1997. *Sissevaateid folkloori lühivormidesse* I [Insights into the Short Forms of Folklore]. Tartu: Tartu Ülikooli Kirjastus.

Krikmann, Arvo 2001. Viimane pikk pilk "Proverbia septentrionalia" valmimisloole [One Last Look at the Compilation of "Proverbia Septentrionalia"]. *Paar sammukest* XVIII. Eesti Kirjandusmuuseumi Aastaraamat. Tartu: Eesti Kirjandusmuuseum.

Õim, Katre 1997. *Eesti võrdluste struktuur* [The Structure of Estonian Similes]. MA thesis. Tartu: The Estonian Language Department of the University of Tartu.

Pall, Valdek (ed.) 1982. *Väike murdesõnastik* 1 [The Small Dialect Dictionary 1]. Tallinn: Valgus.

Pall, Valdek (ed.) 1989. *Väike murdesõnastik* 2 [The Small Dialect Dictionary 2]. Tallinn: Valgus.

Saareste, Andrus 1958. *Eesti keele mõisteline sõnaraamat, II = Dictionnaire analogique de la langue estonienne: Avec un index pourvu des traductions en francais*. Eesti Teadusliku Seltsi Rootsis väljaanne, 3. Stockholm: Vaba Eesti.

Stein, Victor Julius 1875. *Üks kubu Vanu-sõnu ja vanu könekombeid* [A Bundle of Proverbs and Old Speaking Customs]. Tartu: H. Laakmann, 1875.

Vakk, Feliks 1970. *Suured ninad murdsid päid...: (Pea ja selle osad rahvalike ütluste peeglis)* [Big Noses Were Craking Heads... The Head and its Parts in Folk Sayings]. Tallinn: Valgus.

Viks, Ülle 1992. *Väike vormisõnastik* 1: *Sissejuhatus ja Grammatika* [The Form Dictionary 1: Introduction and Grammar]. Tallinn: Eesti Teaduste Akadeemia.