

Word Combinations in the Estonian Dialect Dictionary

Abstract

Estonian dialects are rich in word combinations loaded with meaning, such as phrasal terms, phrasal verbs and idiomatic set expressions. While compiling a dictionary it is necessary to make this material most convenient for the user to find.

This paper shall present the main classes of fixed word combinations used in Estonian dialects and the position and manner of the presentation of different kinds of word combinations in the *Estonian Dialect Dictionary* in progress. The author of this paper is also participating in the research as well as editing process of the compiling of the EDD.

The Dialectological Department of the Institute of the Estonian Language is compiling the first comprehensive dictionary of Estonian dialects (EDD). Dialect material collected for this purpose since the 1920s, has been arranged in a universal alphabetical file kept at the Dialect Archives of the Institute (see also Pall 1994a). At present it contains over 2.7 million file cards. The manuscript has been compiled up to *se*-initial words, edited up to *-ke*, and published from the keywords *a-eksusõit* in three fascicles (EDD 1994, 1995). The dictionary is designed to be published in 8 volumes. For financial reasons it is being published by fascicles.

Parallely the dialectologists and computational linguists have jointly issued the *Väike murdesõnastik* (Concise Dialect Dictionary), which represents an extended list of EDD headwords (VMS 1982, 1989).

The EDD includes all Estonian dialects and the entire dialect vocabulary available. Certain limits have only been set on the presentation of foreign words, newer words in common use, rarer words of children's language and other words not specific to dialect usage. To guarantee the authenticity of the material, only such words have been chosen for publication for which the place, time, the recording person, and in most cases the informant (native speaker of the dialect), can be ascertained. Earlier publications have not been considered in word selection.

EDD headwords are presented in their conventional base form (declinable words in Nom.Sg., verbs in the infinitive in *ma*) and arranged in

the alphabetical order, except the *h*-initial words. Since in most Estonian dialects there is no *h* in word-initial position, so they can be found by the vowel succeeding the initial *h* (see also Oja 1994).

The headword is followed by the head of the entry, containing the phonetic variants that occur in the dialects. Those variants are presented in phonetic transcription and supplied with information on their area of distribution. For words in which the stem is changed in the course of declination or conjugation, both forms of the stem are presented (e.g.: Nom. *hammas* Gen. *hamba*; Infinit. *eitmä* Pres. *eedä*).

The EDD shares certain traits with a translation and defining dictionary, since every dialect word is provided with literary equivalents or explanations. Polysemantic words have every sense or subsense rendered separately.

In order to illustrate and specify the semantic and grammatical aspects of the headword, a number of usage examples from different dialect areas are placed after the literary language explanation, thus taking the user back to the authentic dialect atmosphere.

To a certain extent the EDD also attempts to fulfil the function of a synonym dictionary by referring to those synonyms that happen to share the headword stem, and does so either at the end of an entry or after paragraphs as necessary.

The original principles underlying the compilation of the EDD can be found in the *Prospectus of EDD* edited by Mari Must and designed as a handbook for the compilers of the dictionary (EMSP 1968). When the compilation had reached the letter *l*, however, the accumulated experience prompted some structural changes to be undertaken. Most of the changes concerned the form of the headword, reference system and the way phraseological material was presented. The final structural principles have been formulated by Helmi Neetar in the EDD's Fascicle 1 (Neetar 1994). The author of the present paper, together with Ellen Niit have made a special study of problems with phrases and other word combinations and have worked out suggestions of how to reflect such material in the EDD (Oja 1988).

Estonian dialects are rich in word combinations such as phrasal terms, phrasal verbs and idiomatic set expressions. Considering the fact that a special phraseological dictionary of Estonian dialects could hardly be envisaged, the problem of how to make this material better available for EDD users became essential. The first and perhaps the most important question to ask, is under which headword should a word combination be presented. Consulting the hitherto published alphabetical lists of dialect words, we find that various authors have arranged idiomatic set expressions by their first component. For example, under *k* we find *kurdus*

särk 'a smocked chemise' (Neetar 1976:71) and *känksepä silma suskmine* (a yuletide game) (Keem 1970:480); under *l* there are *lammaskarva hall*, ~ *must*, ~ *valge* 'the colour of natural sheepskin' (Oja 1980:68); while *n* has as an example *nooda joonud* 'slightly drunk' (Neetar 1980:75). Dozens of different names for a roller (*Coracias garrulus garrulus*) *austria vares*, *jaapani vares*, *saksamaa harakas*, *saksamaa pasknäär*, *saksamaa vares*, *samaaria vares* have also been listed by their first component (Mäger 1967:247–262).

Some nature terms consisting of several words have been joined by hyphens, e.g. *maa-alla-mineja* (a fish) (Mäger 1973:33); *südamemajuur* (a herb) (Pall 1994b:215).

Phrasal verbs however, in which the order of the components changes during the course of conjugation, are referred to by means of the more essential component. E.g. the phrase *põldu koguma* 'to gather in crops' is placed alphabetically after *koguma* (Tanning 1961:244) and *põrsod nehveldama* 'to bundle' is under *n* (Neetar 1980:72), whereas *tibasse jääma* (of a woof left out of the web) (Must 1965:406) and *topsu minema* 'to turn out just right' (Sepp 1989:101) can be found under *t* and *ümmargust tegema* 'to go round and round' under *ü* (Must 1995:681).

Here the authors of the EDD proceed directly from the concrete word combination itself, i.e. from the degree of fusion of its components, its degrees of variation, and from its meaning. Our word combinations could thus be conditionally divided into three classes: 1) phrasal terms, 2) phraseologemes, often having a figurative meaning, and 3) phraseological units, in which the components have retained their original meaning.

Depending on the character of the phrase, there are five possible ways to present it in the EDD: 1) as a separate entry, 2) within an entry that is headed by a word occurring in the phrase, either as a) a sub-entry, b) a separate sense group, c) a usage section at the end of a sense group, or d) among case examples of an entry.

1. The most tightly knit units are the so-called phrasal terms. These are word combinations that are used as a common name for a specific object (plant, food, pattern etc.). In the EDD such terms are treated as separate word entries. Usually they are written as a compound word or hyphenated, but a combination of separate words may also occur. The problem of whether they should be treated as compounds or as phrases is merely orthographical and disputable, since the words have been recorded from oral sources. This applies for example to such headwords as *hainlaluselõtõbi* /a nervous disorder in children whose mothers are believed to have been frightened by an adder when pregnant/ (*hainlalune*

is the euphemistic Setu name for an adder); *hallilvatil|vanal|mees* euph. 'wolf'; *heinalmaalkuusk* 'horsetail (Equisetum pratense)'; *pikk-peeter* 'middle finger'; *veskiltiivalkiri* /a textile pattern/ etc. Such phrasal terms in which a word has retained its original meaning or varies, receive the status of a sub-entry under the corresponding headword. E.g. there is an entry *ader* 'plough' with such semantic groups as *harkader* 'forked plough' and *hõlmader* 'mouldboard plough'. *Harkader* in turn includes *läti ader* lit. 'Latvian plough' and *vene ader* lit. 'Russian plough', while *hõlmader* includes *saksamaa ader* lit. 'German plough' and *suur ader* lit. 'big plough'.

Occasionally a phrasal term may include a word that is common to many terms of one kind, e.g. names of illnesses such as *prantsuse haigus* 'syphilis' lit. 'French illness', *äkine haigus* 'stroke' lit. 'sudden illness', *langev haigus* 'epilepsy' lit. 'falling sickness', *jooksja haigus* 'rheumatism' lit. 'travelling illness' etc. Such names may be gathered under one or more sense groups, as may be necessary within the base word entry. The entry *haigus* 'illness' contains a sub-entry *eri haigused* 'various illnesses'. This in turn is divided into 7 subdivisions, the first six of which are dedicated to the names of distinct illnesses (the subdivision for grippe, for example, carries such terms as *hispaania haigus* lit. 'Spanish illness', *inglise haigus* lit. 'English illness', *rahukohtu haigus* lit. 'county court's illness', *vastnõ haigus* lit. 'new illness'. In the seventh and final subdivision however, names are presented which do not fit into a separate subgroup, because of a lack of reliable material to provide an adequate identification of their meaning (e.g. *nurgasängi haigus* (of childbirth), *ummel haigus* 'constipation', *valvaja haigus* 'hernia', *väntind haigus* 'sprain', *väristamise haigus* 'malaria' lit. 'trembling illness' etc.).

If material is scarce, it is not always possible to decide unequivocally whether the word combination in question is an established term or a chance coinage. In such a case the phrase is presented among usage examples within the entries of every one of its essential component words, along with an explanation in parentheses. For example in the Otepää dialect there is a report of a large cross-like pattern used on knitted mittens being called *aoga rehitu* lit. 'slapped with a twig'. But as there are no other reports to corroborate this information, the phrase is presented among the examples to *hagu* as well as to *rehkmä*.

2. There are also very many fixed word combinations that despite having no term value, have an established, often figurative meaning. While occurring separately or in another context, these individual components mean something quite different. Belonging to this group are phraseol-

ogemes, noun-particle combinations with a figurative meaning, noun-verb phrases in which the meaning of at least one component has shifted, etc. Such set expressions are, as a rule, presented as sub-entries under each entry headed by a component (or major component) word of the expression.. E.g. the entry *ahi* 'stove/oven' contains such phraseological sub-entries as *vesi ahjus* 'in for trouble', 'on the verge of tears' lit. 'water in the oven' and *must siga ahjus* 'the fire has gone out in the stove' lit. 'a black pig is in the stove'. The former expression should be entered under *vesi* 'water' as a sub-entry, while the latter should appear both under *siga* 'pig' and *must* 'black'.

If the components of a phraseogeme are subject to variation, the main variants are presented in the sub-entry. Consequently the sub-entry *asi ants ~nudi, ~tahe, ~vask* 'that's that/ problem solved' occurring under *asi* 'thing', contains four phrases (*asi ants, asi nudi, asi tahe, asi vask*), each of which is also presented as a sub-entry in the entry headed by its second component. If there are many variants, all are listed in the sub-entry, while some of them are presented under the sub-entry as usage examples.

Sub-entries that do not fit into any sense group go to the end of the entry as a separate sense group. E.g. *heakene* is the diminutive of the word *hea* 'good', but as the phrase *heakene küll* means 'all right', it gets a separate sense group in the form of a sub-entry.

Such set expressions of which it is difficult to form a sub-entry due to extensive variation in wording, are presented in usage examples together with explanations. Phrases that are similar in meaning are presented in groups separated by vertical lines. If an entry contains many such phrases, they may be gathered under a separate sense group. E.g. the entry *hammas* 'tooth' contains a sense group labelled figurative, which is in turn divided into subgroups. A subgroup contains a number of usage examples that are in turn divided into groups according to their meaning: (of virulence, spiteful speaking, derision) *Hammast päidi alade tõise küliss; naine näüt oma tagumasi hamba; võtab iga ühe seljas oma ammastega; kiristab ambaid meie piale* etc. |(of laughing) *hambad paljalla; ambad laialla; ambad lahti; ambad alati ütsakile* etc. |(of eating) *Tulist suppi saab süia pikka ammastega* (carefully); *sõi nenda pitkade ammastega* (without much appetite) etc.

3. Phraseological unities, all the components of which retain their meaning both as individual words and in other word combinations, usually are not given a separate entry. The user will usually find them in examples placed in the entries headed by every essential word of the phrase. E.g. the phrase *habet ajama* 'to shave' is exemplified both under

habe in the sense group 'beard' and under *ajama* in the sense group 'shave'.

If an entry includes many figurative idioms or phraseological expressions, they may be gathered under a separate sense group labelled figurative. In the case of especially expressive and extensive material, phrasal examples of similar contents are assembled either in a separate sense group or in a special usage section of an appropriate sense group, provided with explanations and examples. Such a sense group entitled "popular sayings and expressions" occurs, for example, under the headword *haukuma* 'to bark'. Here one will find a number of proverbs and sayings that the authors have deemed appropriate to single out, even though the verb *haukuma* is manifested in nothing but its usual meaning.

In exceptional cases a sub-entry could also be evoked by a phrase in which the components function in their original meaning. This may occur if the word is really polysemantic and the dialects' yield of relevant material has been particularly abundant and interesting. Thus the entry *aeg* 'time' includes a sub-entry *ühel ajal* 'simultaneously', 'at the same time'.

Since the degree of fusion occurring in a phrase does not allow for precise measurement, it is difficult to avoid a certain amount of subjectivity on the part of the users and editors of the dictionary. Yet in most cases, the position of a phrase in the dictionary is determined by the material: its volume, its expressiveness, and the meaning of individual words. Every lexicographer knows, that in reality each concrete case should be dealt with individually. However, as both the compilers and the users of a dictionary need the support of certain principles, we have devised the above system to regulate the presentation of word combinations in the *Estonian Dialect Dictionary*.

References

- EDD 1994–1995. *Eesti murrete sõnaraamat I. 1–3*. Tallinn. Eesti Teaduste Akadeemia. Eesti Keele Instituut.
- EMSP 1968. *Eesti murrete sõnaraamatu prospekt*. Tallinn. Eesti NSV Teaduste Akadeemia. Keele ja Kirjanduse Instituut.
- Keem, H. 1970. *Tartu murde tekstid. Eesti murded III*. Tallinn. Eesti NSV Teaduste Akadeemia. Keele ja Kirjanduse Instituut.
- Must, M. 1995. *Keskmurde tekstid. Eesti murded II*. Tallinn. Eesti NSV Teaduste Akadeemia. Keele ja Kirjanduse Instituut.
- Must, M. 1995. *Kirderannikumurde tekstid. Eesti murded V*. Tallinn. Eesti Teaduste Akadeemia. Eesti Keele Instituut.

- Mäger, M. 1967. *Eesti linnunimetused*. Tallinn. Eesti NSV Teaduste Akadeemia. Keele ja Kirjanduse Instituut.
- Mäger, M. 1973. *Juhend ja nimestik kalanimetuste kogumiseks*. Tallinn. Eesti NSV Teaduste Akadeemia. Emakeele Selts.
- Neetar, H. 1976. "Hiiumaa sõnu." In: *Kodumurre 13*. pp. 66–72.
- Neetar, H. 1976. "n-algulisi Hiiumaa sõnu." in: *Kodumurre 14*. pp. 72–78.
- Neetar, H. 1994. "Sõnaraamatu ülesehitus." In: *EMS (1994)* pp. 17–37.
- Oja, V. 1980. "Valgest mustani eesti rahva kõnepruugis." In: *Kodumurre 14*. pp. 68–71.
- Oja, V. 1988. "Sõnauhendid Eesti murrete sõnaraamatus." In: *Emakeele Seltsi Aastaraamat 34*. pp. 13–17.
- Oja, V. 1994. (Review of "Eesti murrete sõnaraamat I, 1. vihik a-aju. Tallinn 1994. 214 S.) in: *Linguistica Uralica XXX*. pp. 220–221.
- Pall, V. 1994a. "Sissejuhatus." in: *EMS (1994)* pp. 5–16.
- Pall, V. 1994b. *Idamurde sõnastik*. Tallinn. Eesti Teaduste Akadeemia. Eesti Keele Instituut.
- Sepp, A. 1989. "Kitsa levikuga murdesõnu Tõstamaalt ja Varblast." in: *Kodumurre 21*. pp. 96–109.
- Tanning, S. 1961. *Mulgi murdetekstid. Eesti murded I*. Tallinn. Eesti NSV Teaduste Akadeemia. Keele ja Kirjanduse Instituut.
- VMS 1982, 1989. *Väike murdesõnastik I, II*. Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut. Toimetanud Valdek Pall. Tallinn. Valgus.