

## DATABASE AND COMPUTER ANALYSIS OF ESTONIAN COLOUR ADJECTIVES

Vilja Oja

*Institute of the Estonian Language.  
Academy of Sciences of Estonia Tallinn, Estonia*

**Abstract:** The existing computer database of the colour names used in Estonian dialects contains about 1,400 different colour adjectives. About eighty percent of the terms are either compound words or lexical units integral in meaning. The Estonian system of colour-denoting compounds and lexical units is an open one: on the principle of analogy new units are constantly produced. Therefore it is vital to know the rules and regularities underlying this mechanism. Computer analysis has helped us to clarify, for example, some semantic relations of seemingly chaotic morphological variation of these word combinations.

**Keywords:** colour terms, compound units, computer database, dialect vocabulary, Estonian, lexicology, linguistics, semantics.

### INTRODUCTION

The present analysis of the Estonian colour lexis draws on the material of the Computer Database of Estonian Colour Names. Colour names are defined as words and phrasal units denoting the tones and hues of both chromatic and achromatic spectra. Mainly dialect vocabulary is examined, but colour terms in literary standard are also taken into consideration. In comparison with the literary language, dialect vocabulary is somewhat more varied, more archaic, and certainly less affected by standardisation.

Scholars have been collecting Estonian dialect vocabulary material for more than seventy years. As a result, the general word files of the Dialect Archives of the Institute of the Estonian Language (Eesti Keele Instituut) contain over 2.7 million file cards. The available dialect collections provide ample opportunity for linguistic research. Comprehensive projects,

however, often seem too time-consuming even to begin. One of the aims of the Estonian colour names project, launched in 1993, was to devise a way to speed up and facilitate the study of Estonian dialect vocabulary with the help of computer technology, while limiting the concrete material examined to colour words. The practical need to specify direct meaning of different colour names often arises in the process of translation (language teaching, study, literature, marketing, etc.).

I am much obliged to the Estonian Science Foundation for providing a grant in support of my project, and to Esko Oja, whose expertise with computers was of immense help.

## 1. COMPUTER DATABASE

In order to provide material for computer analysis of colour names as a semantic class of words, I had to start with the preparation of a special database. Currently the database contains about 1,400 different colour terms. Most of the material comes from the above mentioned Dialect Archives of the Institute of the Estonian Language. This material has been compared to that found in the file of literary Estonian, *Eesti-saksa sõnaraamat* /Estonian-German Dictionary/ by F. J. Wiedemann (1973), *Eesti keele mõisteline sõnaraamat* /Estonian Ideological Dictionary/ by A. Saareste (1958–1968), *Väike murdesõnastik* /Concise Estonian Dialect Dictionary/ edited by V. Pall (1982; 1989), *Eesti kirjakeele seletussõnaraamat* /Defining Dictionary of Literary Estonian/ edited by E. Raiet, et al. (1988–1997), *Õigekeelsussõnaraamat* /Orthographical Dictionary/ edited by R. Kull and E. Raiet (1976), and the dictionaries of synonyms by A. Õim (1991) and P. Saagpakk (1992).

The database records are structured as follows:

*Field 1* of a record contains the headword. This is a dialectal colour name in the nominative and genitive singular forms. It is important to present these two cases since they reveal the possible stem alternations manifested in the process of declension of many Estonian nouns (e.g., nom. *punane*, gen. *punase* 'red'; nom. *valkjas*, gen. *valkja* 'whitish'; nom. *tuhkur* gen. *tuhkru* or *tuhkri* 'grey'; nom. *hall*, gen. *halli* 'grey'). The components of compound words have been separated by a vertical line and provided with labels to mark the part of speech and the way of compounding (e.g., *kirsi|punane* 'red like a cherry' S (= Substantive) gen. + Adj. (= Adjective)).

*Field 2* carries semantic data, including general (basic colour), as well as more specific information (finer shades) (e.g., *kirsi|punane* 'dark red', 'red like a cherry', 'brownish red', 'purplish red').

*Field 3* specifies the sphere of usage of the word, listing the items typically characterised by the word in question (e.g., horse, yarn, cloth, building).

*Field 4* gives one an idea of the geographic distribution of the word, listing the more or less extensive dialect areas in which the word is known to be used (e.g., eP 'northern Estonia', M 'the Mulgi dialect', Ran 'the Rannu subdialect').

The rest of the fields reflect the information attached to the word in the above mentioned dictionaries.

In addition to making it easy to find all information available on any colour term, the database enables the user to produce an alphabetic list of the entry words, the same in reverse alphabetical order, a structural selection (e.g., compounds S nom. + cAdj. (= colour adjective); all compound units; words in *-kas* etc.), various semantic lists (e.g., all light colours; all red colours), all colour adjectives associated with a certain object (e.g., animals; textiles), colour terms registered in any particular dialect or subdialect, etc.

## 2. COLOUR TERMS

### 2.1. Definition

Studying vocabulary used in 116 different Estonian subdialects we cannot determine general basic colour terms. As the area of present-day Estonia had two literary standards up to the mid-nineteenth century - the North-Estonian literary language and the South-Estonian literary language - it seems sensible to regard the South-Estonian variant as a separate language, especially its eastern variety (the Tartu and Võru dialects), where the older usage persisted longer. For example in North-Estonian and in Mulgi dialect the names for basic colours are similar to literary Estonian: *valge* 'white', *must* 'black', *punane* 'red', *kollane* 'yellow', *roheline* 'green', *sinine* 'blue', *pruun* 'brown', *hall* 'grey'. In south-eastern Estonia the basic colours should include such words as *verev* 'red', *haljas* 'green', *hahk* 'grey' and *vahane* 'yellow', which are used to denote the neutral tone of the respective colours, irrespective of the object of description. Etymologies for the names of basic colours are described by Mauno Koski (1983) and Vilja Oja (1995a). The names of basic colours and literary basic colour terms really make up a very small part of all colour terms used in Estonian.

Estonian dialects are rich in words denoting various mixed tones and certain specific shades such as *ruuge* 'yellowish red', *ruske* 'reddish brown', 'russet', *varesemuna|roheline* 'bluish green', *rebu* 1) 'reddish yellow', 2) 'grey', *harm* 'light grey', etc. Some older colour terms have a more restricted usage. For example such words as *hiiru* 'grey', *kimmel* 'yellowish grey' or 'brownish grey', *kõrb* 'dark brown', *leet* 'reddish', *lepp* 'alder-bark coloured', *paat* 'yellowish', 'cream-coloured', *raudjas* 'chestnut', *tuhi* 'grey', *tuhkur* 'ash-coloured', *võik* 'cream-coloured' or 'light bay' mostly apply to animals, especially horses.

Colours are characterised first and foremost along three dimensions, viz., tone, or the place of the colour on a spectrum (e.g., *sinine* 'blue', *rohekas|sinine* 'greenish blue', *sinakas|roheline* 'bluish green', *roheline* 'green'); saturation, or intensity (e.g., *ere|kollane* 'bright yellow', *kahvatu|kollane* 'pale yellow'); and lightness/darkness, or figuratively, the proportion of black and white in the colour (e.g., *hele|pruun* 'light brown', *tume|pruun* 'dark brown') These basic attributes can be complemented by specific nuances such as transparency, softness, lustre, etc. In addition, an Estonian colour term may emphasise some other specific quality of the colour referred to (e.g., *õhk|valge* 'airy white', *vesi|hall* 'watery grey', *kuld|kollane* 'golden yellow'). Therefore the notion of a colour term cannot be confined only to words denoting colours on

chromatic and achromatic scales; instead, colour terms include a much broader selection of words.

## 2.2. Motivations

Many Estonian colour names are motivated by a comparison with the colour of a well-known object. This applies to some basic colours (e.g., *roheline* 'green' < *rohi* 'grass', *verev* 'red' < *veri* 'blood') and to a great number of various tones and hues (e.g., *lepp* < *S lepp* 'alder-bark'; *kirsikas* < *S kirss* 'cherry'). The words *kuld* 'gold', *hõbe* 'silver' and *vask* 'copper', when used in the function of colour names could also be listed under this category (c.f. Oja, 1995b). Sometimes comparison is made with an object whose colour can vary. However, a colour name based on a comparison signifies a single, definite colour tone (e.g., *taeva|karva* 'blue' < *taevas* 'sky', *tuhi* 'grey' < *tuhk* 'ash', *sireli|värvi* 'purple' < *sirel* 'lilac', *liiva|kollane* 'beige' < *liiv* 'sand', *kirsi|punane* 'dark red'). Often the degree of lightness, darkness or intensity is expressed in an analogous way (e.g., *süsi|must* 'charcoal-black' < *süsi* 'charcoal' + *must* 'black', *luik|valge* 'white as a swan' < *luik* 'swan' + *valge* 'white', *tuli|punane* 'bright red' lit. 'fiery red') (c.f. Oja, 1995c).

Many colour names have their origin in the colouring substance: its place of origin (e.g., *pariisi|sinine* < *Pariis* 'Paris' + *sinine* 'blue', *orleani|kollane* < *Orlean* + *kollane* 'yellow'), its plant of origin (e.g., *madarane*, *madara|punane*, *marana|verev* 'yellowish red as dyed with bedstraw roots' < *madar* 'bedstraw', *kaselehe|kollane* 'yellow, dyed with birch leaves'), its way of production (e.g., *poti|sinine* 'indigo blue' < *pott* 'chamber pot' (used for collecting one of the main ingredients of the dye)); sometimes the colour name may also refer to the commercial form of the dyeing substance (e.g., *pakimust* < *pakk* 'packet' + *must* 'black'), etc. In most cases these terms are largely associated with certain specific shades of colour. For example, *orleani|punane* 'Orleans red' means 'yellowish red', the term *liina|punane* (< *aniliin* 'aniline') indicates a purplish red, while *madara|punane* (< *madar* 'bedstraw') refers to a brownish red tone. How adequately such information is received in the process of communication depends on how well-known a given term happens to be.

Borrowed foreign terms are often adapted to fit the phonetic and lexical system of the dialect into which they are borrowed. Generally several variants of such loans are found. For example, *orleani|kollane* ~ *orleani|punane* coexist with *ordeni|kollane* ~ *ordjon* ~ *ordijoon* ~ *ordijaan*, or simply *ort* or *ordi|karva*. The colour terms *liina*, *liine*, *hani|lilla*, *hanililli|sinine*, *aniliine|lilla* etc. indicate an aniline colour prepared from phenylenediamine. The term *pordu|punane* refers to something the colour of Bordeaux wine; literary Estonian has the terms *bordoo* or *bordoo|punane* 'red as claret' < *bordoo* 'claret' + *punane* 'red'. Germ. *neublau* 'new blue' is rendered in dialects spoken on the Estonian island of Hiiumaa as *neiblu|sinine*, *neibla|sinine*, *neibu*, *nein|puu* or *leinpui|sinine*. Purple or violet (Estonian lit. *violett*) in dialect speech is known as *viuli* (~ *viulite*) *värv* lit. 'the colour of violin' or *viuli|punane* 'violin-red' and *viuli|sinine* 'violin-blue'.

Any aspect characteristic of a colour can be emphasized by supplementing the colour word with an adjectival attributive component (e.g., *punakas|kollane* 'reddish yellow', *erk|roheline* 'bright green', *vana|roosa* 'specific shade of pink' lit. 'old pink', *valss|hahk* 'whitish grey', *kütse kõrb* 'rich brown', *hale|sinine* 'pale blue') (c.f. Oja, 1996). One way to describe a colour

is by attaching a diminutive suffix or word to the colour adjective (*sinakas* ‘blueish’, *paatjanõ* ‘yellowish’, *musta|juhti* ‘blackish’, *halli|natu* ‘greyish’, *vereva|poolne* ‘reddish’, *haha|sugumanõ* ‘greyish’, *pool|kollakas* ‘half-yellowish’, *rohe|eliline* ‘greenish’, etc.).

### 2.3. Structure

In Estonian dialects colours can be expressed by root words, adjectival suffixed derivatives, and phrasal units. For the purpose of the present study the first two groups need not be differentiated, since from the semantic point of view the presence or absence of a suffix is not as important to the meaning of the adjective in question as is the influence of a concrete suffix.

Both in literary Estonian and in dialect usage denominal adjectives are often produced by adding a neutral adjectival suffix (*-ne*, *-line*) to a noun stem. In precisely this way have such names for basic colours emerged as *punane* < *puna*+*ne*, *roheline* (*rohiline*, *rohilane*) < *rohi*+*line* (*-lane*), *sinine*, *kollane* (*koldne*, *kõllane*, *keldane*), *vahanõ* (in Võru dialect). The same applies to metallic colour derivatives *kuldne*, *kullane* ‘golden’, *hõbedane* ‘silvery’, *vaskne* ‘coppery’, and to several other colour adjectives such as *roostene* ‘rust-brown’, *purpurne* ‘purple-red’, ‘crimson’, *kahjane* ‘vat black’, *linune* ‘linen white’, etc.

The meaning of adjectives with the suffixes *-kas* and *-jas*, which occur frequently among colour names, largely depends on the part of speech of the derivation base. If the suffix *-kas* or *-jas* has been attached to an adjective, then the resulting colour name has a meaning that is close to that of the derivation base, but denoting a weaker shade of the same colour (e.g., *punakas* ‘reddish’ < Adj. *punane* ‘red’, *valkjas* ‘whitish’ < Adj. *valge* ‘white’). In the case of denominal adjectives, however, the comparison is mostly based on outward similarity (e.g., *savikas* ‘clay-coloured’ < S *savi* ‘clay’; *raudjas* ‘chestnut’ lit. ‘iron-coloured’ < S *raud* ‘iron’).

In South-Estonian dialects *-jas*- and *-kas* derivatives are used alongside colour adjectives in *-jane* or *-kane* (e.g., *haabjanõ* ‘yellowish grey’, *kõllakanõ* ‘yellowish’, *mustjanõ* ‘blackish’, ‘dark’). The occurrence of derivatives in *-lik* or *-ik* is less frequent (e.g., *hiirik* ‘grey’, *valgõlik* ‘whitish’). Colours may also be formed by participles in *-nud* or *-v* (e.g., *sinav* ‘appearing blue’, *kulõhunu* ‘greyish’).

Most of the colour names available in the database are either compound words or phrasal units. All compound colour names can be analysed into two components, viz., the attributive (first) part and the base (final) component. Either of these components may occur separately as a root word, suffixed derivative, compound word, or word combination. As colour words are defined by meaning, the present analysis of compound colour names is also mainly based on the semantic aspect. Hence the necessity to single out the class of colour adjectives (cAdj.) from other adjectives (Adj.). This approach yields the following classification, according to which the Estonian compound colour words fall into nine categories:

1) cAdj gen./part. + parameter word (e.g., *punast karva*, *punase karvaga*, *punasekarvaline* ‘red’). The addition of the word *karv* to a nominal or genitive attribute has been considered an ancient Finnic way of colour description. In Estonian dialects we find other analogous parameter words such as *värv*, *nägu*, *toon*, and *lee*. Compound terms whose base word is an

adjectival derivative of a parameter word (e.g., *-karvaline*, *-värviline*, *-tooniline*, *-näoline*) should be regarded as derivative compounds made up of a compound word and a suffix, or as compound derivatives. As for compound colour terms, the addition to them of a suffix effects neither the terms' meaning nor their sphere of application. The words *halli karva* = *hallikarvaline* = *hallikarvane* = *halli karvaga* = *halli nägu* = *hallinäoline* = *hall* 'grey' are all practically interchangeable.

2) S gen. + parameter word (e.g., *vere|karva*, *vere|karvaline* 'blood-coloured'). If the attributive component is a noun, it functions as the base of reference, denoting an object the colour of which is generally known. For example, *maksa|karva* 'dark brown' lit. 'liver-coloured', *merekarva* 'greyish blue or green' lit. 'of the colour of the sea', *küpse kirsi värvi* 'dark red' lit. 'of the colour of a ripe cherry', *tolmu|karvaga* 'grey' lit. 'of the colour of dust', *marja|näoline* 'pink' lit. 'berry-coloured'. In the case of dyed cloth the attributive component may denote the dyeing substance, e.g., *sambla|karva* 'brown (dyed with moss)', *orase|karva* 'greenish grey (dyed with green sprouts)'.

3) S nom. + parameter word (e.g., *lammas|karva* 'sheep-coloured'). This compound denotes the colour of natural (uncoloured) lambswool. The colour term *lammas|karva* is used to talk about natural white, black and grey lambswool, as well as woollen yarn (or textiles) of that colour.

4) cAdj. gen./part. + *-juhti*, *-vuhti*, *-võitu*, *-krunti*, *-natu*, *-poolne*, *-sugune*. (e.g., *punast juhti*, *punase|võitu* 'reddish'). The bound final component *-juhti*, *-võitu* or some other such word specifies the degree of manifestation of the quality in question.

5) S gen. + *-juhti*, *-vuhti* (e.g., *vere juhti* 'redish' lit. 'bloodish'). The attributive noun denotes an object whose colour is generally known. The compound adjectival colour term describes a colour similar to that of this object.

6) cAdj. nom. + cAdj. (e.g., *pruunikas|punane* 'brownish red'). If the base word is a colour adjective, it may take either an adjective or a noun as the attributive part. If the attributive part is an adjective, it usually serves to specify one of the basic qualities of the base word colour, i.e. its tone, lightness/darkness or intensity. The tone is mostly specified by another colour name in adjectival function. Such a compound of two colour names describes mixed hues.

7) Adj. nom. + cAdj. (e.g., *tumepunane* 'dark red'). The number of adjectives used to specify degrees of lightness/darkness or intensity is relatively small. (See Oja, 1996) In some cases the function of the adjective is performed by an adverb of degree. (See Oja 1997b:29)

8) S gen. + cAdj. (e.g., *kirsipunane* 'red like a cherry'). Many of these are based on comparison, as the attributive word is a noun denoting the base of reference. As with the previously discussed compounds of a noun and a parameter word, the base of reference underlying these comparisons may be any object familiar to the speakers. Another group of the compounds have come from colouring substance (e.g., *koosenikollane* 'in-grain yellow' < *košenill* 'grain' made of cochineal insects) (see 2.2.).

9) S nom. + cAdj. ((e.g., *tulipunane* 'fiery red'). In dialect speech such nominal attributive word serves to emphasise the intensity of the colour. In literary Estonian, colour name categories 8 and 9 partially mixed (Rannut, 1966).

In the categories 1 through 5 the colour is generally emphasised by the attributive part of each compound unite; however in the categories 6 through 9 the final component describes the base colour. Most of the analysed units fall under category 6 or 9 (Oja, 1997a:15–18). As demonstrated by a more detailed analysis, the difference in combinations, originating in the part of speech of the compounds and the way of compounding, may be semantically significant. For example, in Estonian dialects the S nom. + cAdj. combination denotes mainly the intensity of the colour, whereas S gen. + cAdj. is mostly used to specify the shade of the colour (Oja, 1995c).

#### 2.4. Comparison with contact languages

Estonians have had long-lasting historical, cultural and linguistic contacts, on the one hand with Finnic nations, and on the other, with Germans. The system of Estonian colour words in general shares some traits with other Finnic languages and some traits with the respective German system. Like Estonian, both Finnish and German have colour names with a modifying suffix (e.g., Germ. *grünlich*, Est. *rohekas*, Fin. *vihreähkö* 'greenish'). A great number of compound colour names in German and Finnish have the same structure as in Estonian (e.g., Germ. *schneeweiss*, Est. *lumivalge*, Fin. *lumivalkoinen* 'snowy white'; Germ. *dunkelbraun*, Est. *tumepruun*, Fin. *tummanruskea* 'dark brown'; Germ. *goldfarbig*, Est. *kullakarva*, Fin. *kullanvärinen* 'golden' etc.) (cf.: Oksaar, 1961; Koski, 1983; Oja, 1995a; Oja, 1997b).

On occasion the three languages may even confront the scholar with similar problems. The identification of the part of speech of a colour name is bound to give more trouble in German than elsewhere, as part of speech in German has a direct bearing on capitalisation. Another source of confusion is the variation of nominal and genitive compounding. Several colour terms apply only to a restricted class of objects. Also, the meaning of a term often requires additional specification by indicating its position on the chromatic (colour) and/or the achromatic (black-and-white) scales. It seems rather important that in all these three languages colour names belong to an open system. This means that their number is not limited; proceeding from the structures governing the language, the speaker may feel free to produce his or her own colour names. If one wishes to be acceptably creative in the use, maintenance or development of a verbal language, one has to be competent in the regularities underlying the functioning of the language.

#### REFERENCES

- Koski, M. (1983). *Värien nimitykset suomessa ja lähisukukielissä*. Suomalaisen kirjallisuuden seuran toimituksia 391. Savonlinna.
- Kull, R. and E. Raiet (1976). *Õigekeelsussõnaraamat*. (R. Kull and E. Raiet (Ed.)). Valgus, Tallinn.

- Oja, V. (1995a). Colour Terms in the Finnic Atlas. In: *Itämerensuomalaisen kielikartaston symposiumi. FU 8 Jyväskylä 12. 8. 1995.* pp. 39–66. Kotimaisten kielten tutkimuskeskus, Helsinki.
- Oja, V. (1995b). Gold, silver and copper in the Estonian colour names. *Linguistica Uralica* XXXI, 3, pp. 178–187.
- Oja, V. (1995c). Võrdlusel baseeruvaist värviliitsõnadest. *Keel ja Kirjandus* 12, pp. 809–817.
- Oja, V. (1996). Compound adjectives used in the Estonian dialects to denote light and pale shades of colour. *Linguistica Uralica* XXXII, 2, pp. 92–100.
- Oja, V. (1997a). *The Designation of Colours in Estonian Dialects*. Tallinn Pedagogical University, Tallinn.
- Oja, V. (1997b). *Värvinimetuste moodustamine eesti murretes*. Preprint. Eesti Teaduste Akadeemia. Eesti Keele Instituut, Tallinn.
- Oksaar, E. (1961). Die Farbenbezeichnungen im Deutsch der Gegenwart. *Muttersprache*, Juli/August, pp. 207–220.
- Pall, V. (1982; 1989). *Väike murdesõnastik* I; II. (V. Pall (Ed.)). Valgus, Tallinn.
- Raiet, E., et al. (1988–1997). *Eesti kirjakeele seletussõnaraamat* I – IV. (E. Raiet, et al. (Ed.)). Eesti Teaduste Akadeemia. Eesti Keele Instituut, Tallinn.
- Rannut, L. (1966). Nimisõna liitumine värvust tähistava omadussõnaga. *Keel ja Kirjandus* 8, pp. 478–486.
- Saagpakk, P. F. (1992). *Sünonüümi-sõnastik*. Maarjamaa, Brampton.
- Saareste, A. (1958–1968). *Eesti keele mõisteline sõnaraamat*. Vaba Maa, Stockholm.
- Wiedemann, F. J. (1973). *Eesti-saksa sõnaraamat*. Neljas, muutmata trükk teisest, Jakob Hurda redigeeritud väljaandest. Valgus, Tallinn.
- Õim, A. (1991). *Sünonüümisõnastik*. Oma kulu ja kirjadega, Tallinn.