Systematic Polysemy
of Estonian Colour Adjectives

Maria Tuulik[1, 2], Margit Langemets[1]

Institute of the Estonian Language[1], University of Tartu[2]
e-mail: Maria.Tuulik@eki.ee, Margit.Langemets@eki.ee

Abstract

This article introduces patterns of systematic adjectival polysemy in Estonian focusing on colour terms. The aim of our work is to create a theoretical framework for presenting adjectives in the Estonian Collocations Dictionary (to be published in 2018). To explore systematic polysemy patterns, semantic types were attached to the word senses and similarities in the semantic shifts of the words were searched for. Semantic types provided for describing Estonian (polysemous) adjectives (Tuulik 2014) were used in the study. In the case of senses denoting nouns, semantic types developed for Estonian nouns (Langemets 2010) were used.

The colour adjectives are examined in terms of both polysemy and morphological structure. A polysemy index, i.e. the average number of meanings per word, is ascertained on the basis of a single-volume explanatory dictionary of Estonian (to be published in 2018) for root adjectives, derivatives and compound words in order to explore the relations between polysemy and morphology based on the example of Estonian colour adjectives (463 colour terms altogether).

Keywords: semantics; lexicology; lexicography; morphology; adjectives; Estonian

1 Introduction

While discussing semantics of adjective, one always faces the problem to what extent the meanings of an adjective can be independent or context-bound. One of the characteristic features of adjectives is that depending on the context, an adjective may refer to different features of the noun it modifies. Thus the meaning of an adjective may change depending, on the one hand, on the noun's semantic component referred to, and, on the other hand, on the type of the noun it modifies (see, e.g. Bhat 1994). For example, the yellow tones in "yellow lemon", "yellow moon" and "yellow skin" are obviously quite different in terms of a colour map. The present study addresses adjective semantics from the perspective of adjectives, while 'sense' refers to the lexical or dictionary sense of the adjective, leaving aside how this lexical sense should be described.

The Estonian Collocations Dictionary (ECD), for which the colour words are analysed, is a monolingual online scholarly dictionary intended for learners of Estonian as a foreign or second language at the upper intermediate and advanced levels (see Kallas et al. 2015a). The dictionary contains about 10,000 headwords. The collocates within each headword are grouped according to the lexico-grammatical structure formed by the collocational phrase, and example sentences are provided. For the automatic generation of the ECD database, there were used the following functions of the corpus query system Sketch Engine (Kilgarriff et al. 2004) – Word List, Word Sketch and Good Dictionary

Example (GDEX).

Polysemy was analysed semasiologically on the basis of lexicographical descriptions in the ongoing project of the Estonian explanatory dictionary (to be published in 2018/2019, ca 100,000 headwords, Langemets et al. 2010, henceforth EstDic), where word senses are provided with semantic type labels, as well as the ECD database which contains automatically collected corpus data. In addition, the corpus material (mainly the Adj+Noun collocations) is accessed through word sketches. The sample of colour adjectives (collected from EstDic) consists of words with at least one sense labelled as QUALITY_COLOUR. As of April 2016, the total number of such words in the dictionary working base was 463.

2   Polysemy Index of Estonian Colour Adjectives

Based on morphological structure, the 463 colour adjectives can be classified as follows: 45 root adjectives, 348 compounds and 70 derivatives (see Table 1). Words were classified according to the Dictionary of Estonian Word Families (Vare 2012). In some cases the same colour was represented in all three forms, as a root word (for example *lasuur* 'azure')*,* as a derivative (*lasuurne*) and as a compound (*lasuursinine*).

Colour adjectives (EstDic)	Number (Percentage)	Examples
root adjectives	45 (10%)	*kirju* 'multicoloured', *oranž* 'orange', *must* 'black', lilla 'purple'
compounds	348 (75%)	*hõbevalge* 'silver white', *jääsinine* 'ice blue', *maksakarva* 'liver colour', *neoonroheline* 'neon green'
derivatives	70 (15%)	*pronksjas* 'bronzy', *kahvatu* 'pale', *kreemikas* 'creamy', *kollane* 'yellow'

Table 1: Colour adjectives classified by morphological structure.

A vast majority of the analysed colour adjectives are compounds. This agrees with the statement of Estonian word formation expert Reet Kasik, that most lexemes expressing properties in Estonian are complex adjectives – compounds and derivatives (Kasik 2015: 315). A similar morphological division is valid for Estonian nouns (Langemets 2010: 86-87). In addition, the large proportion of compounds in the sample can be partly on account of the peculiarity of colour terms being often based on comparison, for example, *hiirhall* 'mouse grey', *korallpunane* 'coral red' or *kriitvalge* 'chalk white'.

Our sample of colour terms features a great number of relatively rare words, which can be predictive of a lower proportion of root words compared with compound words and derivatives. Indeed, a comparison of the morphological structure of those colour adjectives with that of the 100 most frequent Estonian adjectives (Tuulik 2014) revealed that the percentage of root adjectives was 34% for frequent adjectives compared with just 10% for colour terms. Moreover, taking a separate look at basic colour terms (for details on Estonian basic colour terms see Sutrop 1995: 797–808), whose frequency is high, we will also see that the proportion of root words is overwhelming, notably, as many as 7 out of the 11 basic colour terms are root adjectives, while the remaining 4 are

derivatives. This conforms to Zipf's Law (principle of least effort) (1949), according to which the most frequent words are also the shortest. All types of derivatives present in the sample are shown in Table 2.

Derivatives	Number	Examples
ne-derivatives	19	*punane* 'red', *kuldne* 'golden', *smaragdne* 'emeraldy', *suitsune* 'smoky'
jas-derivatives	15	*tuhkjas* 'ashy', *valkjas* 'whitish', *hiirjas* 'mousy'
kas-derivatives	13	*pruunikas* 'brownish', *roosakas* 'pinkish', *sinakas* 'bluish'
line-derivatives	8	*roheline* 'green', *roosiline* 'rosy', *värviline* 'colourful'
others (*tu*-, *lik*-, *as*- derivatives)	15	*värvitu* 'colourless', *akvarellilik* 'watercolour-like', *haljas* 'verdant'

Table 2: Derivatives classified by suffixes.

The number of senses comes from the finished part of EstDic (to be published in 2018). In the study we did not distinguish between core sense and subsense, all units were counted and analysed. Analysis of the sample by the number of semantic units resulted in Table 3. Most of the sample words (348 words, i.e. 75%) were monosemous, while only 115 words (25% of all colour adjectives) had more than one dictionary sense. The most polysemous words in the sample were *must* 'black' (10 semantic units) and *valge* 'white' (9 semantic units).

Number of sense units	1	2	3	4	5	6	7	8	9	10
Number of adjectives	348	75	19	10	5	2	1	1	1	1

Table 3: Adjectives based on their number of sense units.

To explore the relations between polysemy and morphological structure, a polysemy index, i.e. the average number of meanings per word was ascertained for root adjectives, derivatives and compound words (Table 4). The table reveals that root word polysemy is about twice higher for root words than for derivatives or compound words, while compounds hold the monosemy record of the sample (perhaps partly due to their length). The general linguistic tendency of root words being more polysemous than the rest is also there for colour adjectives. Langemets (2010: 271) points out that systematic polysemy – as polysemy in general – is more likely to occur with a simple word than with a compound one, thus preferring shorter (and more synthetic) units over longer (and more analytic) ones.

Colour adjectives	Polysemy index
Root adjectives	3.2
Derivatives	1.7
Compounds	1.2

Table 4: Polysemy index for colour adjectives.

3 Polysemy Patterns of Estonian Colour Adjectives

To explore systematic polysemy patterns, semantic types were added to the word senses and similarities in their semantic shifts were searched for. Semantic types provided for describing Estonian (polysemous) adjectives (Tuulik 2014) were used in the study. In the case of senses denoting nouns, there were used semantic types developed for Estonian nouns (Langemets 2010).

Although some regularities were observed in our semantic studies, it does not tell us where systematic polysemy begins. So we followed Apresjan's definition as summarized by Langemets (2010: 265): "Systematic polysemy is a situation where several senses of at least two words regularly imply a similar semantic structure".

The semantic structures or patterns were derived from all 115 polysemous colour words available in the sample. Table 5 provides a survey of the regular semantic alternations observed in Estonian colour words.

With colour words, the most systematic universal semantic alternation is QUALITY_COLOUR (adjective) – COLOUR (noun); in principle, any colour adjective can be subjected to such nominal use. In some cases it can be difficult to decide whether one has to do with a noun or an elliptical use of an adjective, where the word for 'colour' has been omitted. Despite the pattern being universal theoretically, it is hardly salient in DicEst. The most productive patterns for this dictionary were COLOUR–MATERIAL/SUBSTANCE and QUALITY_COLOUR–QUALITY_MATERIAL/SUBSTANCE.

Pattern	No. of words	Examples	Pattern description
QUALITY_COLOUR–COLOUR (ADJ+N)	Possible for all colour terms	*kollane* 'yellow', *roheline* 'green', *punane* 'red'	The colour adjective also occurs in nominal use (same hue), e.g. "some yellow in the birches", "wearing red"
COLOUR–MATERIAL/SUBSTANCE (N+N)	20	*vask* 'copper', *neoon* 'neon', *mahagon* 'mahagony'	The noun stands for a material or substance as well as for the corresponding hue (e.g. *copper* as a metal and a reddish hue)
QUALITY_COLOUR–QUALITY_MATERIAL/SUBSTANCE (ADJ+ADJ)	16	*pigine<pigi* 'pitch', *tindine<tint* 'ink', *süsine<* 'coal' 'süsi*, *pronksine*' 'bronze' Adj.	The adjective refers to an object being (all or partly) covered with a substance or material, or being made of it, as well as being of the corresponding colour. (E.g. *süsine* can mean "covered with coal; burnt to coal" or "absolutely black, extremely dark".)
QUALITY_COLOUR–QUALITY_ASSESSMENT (ADJ+ADJ)	13	*hall* 'grey', *kahvatu* 'pale', *roosiline* 'rosy'	The colour sense alternates with that of assessment (e.g. grey meaning "dull" and pale referring to "unimpressive")
QUALITY_COLOUR–HUMAN_QUALITY (ADJ+N)	11	*blond* 'blonde', *brünett* 'brunette', must 'black', valge 'white', värviline 'coloured'	The word can refer to a colour as well as to a person (noun) of this or that quality. (E.g. words of hair colour such as *blond* 'blonde' and *brünett* 'brunette' can also denote a person with hair of that colour.)
QUALITY_COLOUR–ANIMAL_QUALITY (ADJ+N)	9	*kõrb* 'sorrel, chestnut, bay',	The word can refer to a colour as well as to a horse (noun) of that colour (e.g.

		hiir 'dun', võik 'light bay'	kõrb is "reddish brown" and a "reddish brown horse")
QUALITY_COLOUR–QUALITY_PSYCH (ADJ+ADJ)	9	*must* 'black', *roosa* 'pink', *terasene 'steel'*, *erk* 'bright'	The colour sense alternates with characteristics of mood, personal nature or mental ability, e.g. *black* as depressing or pink as "cheerful, carefree".
QUALITY_COLOUR–QUALITY_VOICE (ADJ+ADJ)	9	*hele* 'light', *tuhm* 'dull', *hõbedane* 'silvery', *matt* 'toneless'	The colour sense alternates with a quality of a voice or sound, e.g. *silvery* also meaning "high, sounding brightly".
QUALITY_COLOUR–QUALITY_SOCIETY–IN_ROLE (ADJ+ADJ+N)	6	*punane* 'red', roheline 'green', *valge* 'white', *roosa* 'pink', *punaroheline* 'red and green'	A colour associated with a political movement or way of societal thinking often sticks to that movement or way of thinking as well as to a proponent of the ideology (e.g. *red* as a communist and *green* as a proponent of ecological ideology).

Table 5: Overview of polysemy patterns for colour terms (based on the DicEst sample).

4 Presentation of the Polysemy Patterns in the ECD

The way and regularity of presentation of polysemy patterns in dictionaries depend on the purpose and target group of the dictionary. Our aim was to develop such principles for the Estonian Collocations Dictionary (see Introduction). The list of the patterns is based on a large monolingual general-language dictionary of Estonian, which contains ca 100,000 headwords. While the ECD is a learners dictionary of ca 10,000 headwords. The list of headwords as well as the database together with collocation candidates and examples are all generated automatically, based on frequency, from the Estonian National Corpus (463 million words) using the Sketch Engine (for database details see Kallas et al. 2015b). About 1500 of 10,000 ECD headwords are adjectives. The last group of colour senses will be included in the dictionary by the time of its completion.

The patterns are based on a far greater number of colour terms than included in the ECD list of headwords. We attempted to determine which patterns are vital for the ECD and how they should be presented. The general principles are as follows: If both senses of a colour adjective are manifest in the CDE database, they will both be made explicit in the CDE. If only one sense of an adjective belonging to a pattern is manifest in the CDE database, then word sketches are examined directly in the corpus to make the decision about presenting or omitting the other sense.

Also in the case of the QUALITY_COLOUR–HUMAN_QUALITY pattern, all sample words were present in the ECD headword list and both senses of each were frequent in the database, so they will all qualify for the ECD. For the sake of clarity, in the presentation of collocations the senses taking different parts of speech are given separate headwords. Another pattern with both senses occurring in the database, as well as in sketches, was QUALITY_COLOUR–QUALITY_VOICE (e.g. *hele* 'bright', *tuhm* 'dull', *hõbedane* 'silvery'), and so it was decided to systematically present both senses in the dictionary. The N+N pattern COLOUR–MATERIAL/SUBSTANCE would mostly cover low-frequency words, such as "cobalt blue" or "cadmium red", which are not part of the ECD headword

list. The more salient (frequent) representatives of the pattern, such as *rooste* 'rust', *vask* 'copper', or *pronks* 'bronze', were included in the ECD without the colour sense, as the latter was rare enough in the corpus not to appear in word sketches. Even though there was some semantic alternation, it concerned noun senses and as such it is beyond our topic of systematic polysemy of adjectives. Although most of the words representing the pattern QUALITY_COLOUR–QUALITY_MATERIAL/SUBSTANCE also belonged to the lower end of the frequency scale, for those words that were contained in the ECD list (e.g. *kuldne* 'golden', *hõbedane* 'silvery') both senses could be detected in the database, which meant systematic inclusion of both senses in the dictionary.

Both senses are also eligible for the inclusion in the ECD from the group of words following the pattern QUALITY_COLOUR–QUALITY_ASSESSMENT as most of the assessment senses (e.g. for *hall* 'grey', *kahvatu* 'pale', *kuldne* 'golden') were evident in the (frequent) database collocations.

Systematic presentation was abandoned in the case of two patterns, namely, QUALITY_COLOUR–QUALITY_PSYCH and QUALITY_COLOUR–QUALITY_SOCIETY–IN_ROLE, where both senses will be included only in case of higher frequencies of occurrence. For example, collocations will be presented for black as "depressing" and pink as "careless, optimistic", but not for green in the sense of "naive". Also, political senses are included for green and red, but not, e.g., for pink and white.

Such ADJ+N patterns as QUALITY_COLOUR–COLOUR and QUALITY_COLOUR–ANIMAL_QUALITY were left aside, because the former had no nominal collocations in the database and the latter is represented by low-frequency words, which did not make their way into the ECD headword list.

5 Conclusion

In the present study, 463 Estonian colour terms were described and analyzed with regard to both their polysemy and morphological structure. There was also examined the polysemy index, i.e. the average number of meanings per word, for root adjectives, derivatives and compound words. The most polysemous of these are root adjectives, with a polysemy index of 3.2. The polysemy index of derivatives is about a half lower: 1.7, while the least polysemous (polysemy index 1.2) in the sample were compound words. The analysis of systematic polysemy patterns of Estonian colour adjectives was based on their descriptions in the explanatory dictionary of Estonian (EstDic) and on corpus data. In order to ascertain systematic shifts in meaning, there were used semantic types. Nine patterns applying to colour words were revealed. We also investigated which of those patterns were more salient in the corpus material and how to present them in the collocation dictionary (ECD).

The present article surveyed the systematic semantic alternation of Estonian colour words. In the future, other semantic groups of adjectives will be addressed. The long-term goal is to map the polysemy of the Estonian adjective as a whole, developing a theoretical framework for a systematic presentation of adjectival polysemy in dictionaries.

6 References

Bhat, Shankara D. N. (1994). *The adjectival category: criteria for differentiation and identification.* Amsterdam/Philadelphia: John Benjamins.

Dictionary of Estonian (DicEst). To be published in 2018. Manuscript is in the database of Institute of the Estonian Language.

Estonian Collocations Dictionary (ECD). To be published in 2018. Manuscript is in the database of Institute of the Estonian Language.

Kallas, J., Koppel, K., Tuulik, M. (2015a). Korpusleksikograafia uued võimalused eesti keele kollokatsioonisõnastiku näitel. [New Possibilities in Corpus Lexicography based on the Example of the Estonian Collocations Dictionary.] In *Eesti Rakenduslingvistika Ühingu aastaraamat,* 11, pp 75−94.

Kallas, J., Kilgarriff, A., Koppel, K., Kudritski, E., Langemets, M., Michelfeit, J., Tuulik, M., Viks, Ü. (2015b). Automatic generation of the Estonian Collocations Dictionary database. Electronic lexicography in the 21st century: linking lexical data in the digital age. *Proceedings of the eLex 2015 conference*, 11-13 August 2015, Herstmonceux Castle, United Kingdom. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd, pp 1−20.

Kasik, R. (2015). *Sõnamoodustus*. [Word formation.] Tartu: Tartu Ülikooli Kirjastus.

Kilgarriff, A., Rychly, P., Smrž, P. & Tugwell, D. (2004). The Sketch Engine. In G. Williams, S. Vessier (eds.) *Proceedings of the XI Euralex International Congress*. Lorient: Universite de Bretagne Sud, pp. 105-116.

Langemets, M. (2010). *Nimisõna süstemaatiline polüseemia eesti keeles ja selle esitus keelevaras.* [Systematic Polysemy of Nouns in Estonian and its Lexicographic Treatment in Estonian Language Resources.] Tallinn: Eesti Keele Sihtasutus.

Langemets, M., Tiits, M., Valdre T., Voll, P., (2010). In spe: üheköiteline eesti keele sõnaraamat. [In Spe: One-volume Dictionary of Estonian] In *Keel ja Kirjandus*, 11, pp 793–810.

Sutrop, U. (1995). Eesti keele põhivärvinimed. [Estonian Basic Colour Terms.] In *Keel ja Kirjandus*, 12, pp 797–808

Tuulik, M. (2014). Adjektiivide polüseemia korpuses ja sõnaraamatus. [Adjectival polysemy in corpora and dictionaries.] In *Eesti Rakenduslingvistika Ühingu aastaraamat*, 10, pp 307−317

Vare, S. (2012). *Eesti keele sõnapered. Tänapäeva eesti keele sõnavara struktuurianalüüs I–II*. [The Dictionary of Estonian Word Families.] Tallinn: Eesti Keele Sihtasutus.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. AddisonWesley Press.