Folklore

Electronic Journal of Folklore 64

Folklore
Electronic Journal of Folklore
http://www.folklore.ee/folklore
Printed version
Vol. 64
2016

Folk Belief and Media Group
of the Estonian Literary Museum

Estonian Institute of Folklore

Folklore

Electronic Journal of Folklore
Vol. 64

Edited by Mare Kõiva & Andres Kuperjanov
Guest editors: Hille Pajupuu, Renate Pajusalu,
Mari Uusküla

Tartu 2016

Editor in chief	Mare Kõiva
Co-editor	Andres Kuperjanov
Guest editors	Hille Pajupuu, Renate Pajusalu, Mari Uusküla
Copy editor	Tiina Mällo
Design	Andres Kuperjanov
Layout	Diana Kahre

The journal is supported by the Estonian Ministry of Education and Research (IUT 22-5), the European Union through the European Regional Development Fund (Centre of Excellence in Estonian Studies), the state programme project EKKM14-344, and the Estonian Literary Museum.

Indexed in EBSCO Publishing Humanities International Complete, Thomson Reuters Arts & Humanities Citation Index, MLA International Bibliography, Ulrich's Periodicals Directory, Internationale Volkskundliche Bibliographie / International Folklore Bibliography / Bibliographie Internationale d'Ethnologie, Open Folklore, DOAJ, C.E.E.O.L., ERIH Plus, Scopus

Editorial address:
Folklore: Electronic Journal of Folklore
Vanemuise 42–235
51003 Tartu
Estonia
phone: +372 737 7740
fax: +372 737 7706
e-mail: folklore@folklore.ee
home page: http://www.folklore.ee/folklore

ISSN 1406-0957
doi:10.7592/FEJF2016.64

CONTENTS

INTRODUCTION

The 64th issue of *Folklore: Electronic Journal of Folklore* contains researches in the field of linguistics and folkloristics, which are dedicated to categories and their applications essential in terms of culture, map the cognitive backgrounds of language use by different peoples, and highlight the lingvo-folkloric or lingvo-mythological interconnections. Such a selection was substantiated by a need for interdisciplinary confluence, but it was rendered possible by the 60th birthday of Professor Urmas Sutrop, a renowned linguist and cultural researcher. Urmas Sutrop's fields of research have included human perception and cognisance, on the one hand, and the expression of cultural phenomena in language, on the other. He is known on the international scene for the elaboration of the cognitive salience index, which helps to systematise what a group of people regards as important (salient). Urmas Sutrop has made a remarkable contribution to the studies of colour terms, but his interests have also involved other spheres of cognisance, such as temperature, taste, emotions, sign language, language history, and language policy. He has written treatments on theonyms, especially the meaning of the name of Taarapita, as well as on older figures of gods. Urmas Sutrop is editor-in-chief of international journals *Eesti ja soome-ugri keeleteadus / Journal of Estonian and Finno-Ugric Linguistics* and *Trames*, and he has contributed to the publication of a great volume of invaluable cultural riches.

Estonian Jumal *'God'* by Anni Jürine, Karl Pajusalu, Renate Pajusalu, Ilona Tragel, and Ann Veismann focuses on semantics and the use of the word *jumal* 'god' in Estonian. The article gives an overview of the etymology of the word, its meanings, and its role as a linguistic unit fulfilling different communication goals.

The article titled *Identifying Polarity in Different Text Types* by Hille Pajupuu, Rene Altrov, and Jaan Pajupuu describes some options of how to predict the possible effect of a written text on the reader and the creation of an automatic identifier of the polarity (positivity-negativity) of Estonian texts, which is independent of domain and of text type. The Estonian polarity identifier is an open source and available for free.

Man's Gender and Age as Based on the Collocations of the Estonian Word Mees *'Man'* by Liisi Piits explores collocational patterns of the noun *mees* 'man', which reveal social attitudes and stereotypes of age, gender, and behaviour. Differences in the descriptions of men (in comparison to women and boys) as reflected in specific adjectives and verbs are discussed.

In the article *From Listing Data to Semantic Maps: Cross-Linguistic Commonalities in Cognitive Representation of Colour*, Mari Uusküla and David Bimler use non-metric multidimensional scaling of similarities to reconstruct an abstract semantic network. Their analysis suggests that semantic maps are

similar, despite of the typological differences between the languages. All subjects start with salient terms, moving on with prototypal colours, after which they access their semantic knowledge along a trail of associations and start to list less frequent opaque terms, including compounds and derivatives. These findings support the hypothesis that the cognitive organisation of colour terms in our mental lexicon does not follow perceptual similarities.

The article by Jodi L. Sandford, titled *Color Entrenchment in Middle-School English Speakers: Cognitive Salience Index Applied to Color Listing*, presents results of two questionnaires posed to English-speaking middle-school students to verify the level of color term entrenchment and color prototypes at the age of 12.

Colour Symbols in Mari Songs by Natalia Glukhova presents a system of colour symbols in Mari folk songs based on the results of a multifold investigation (semantic, quantitative, statistics, etc). The main reconstructed meanings of colour in the songs denote emotions such as joy, wonder, astonishment, grief, melancholy; aesthetic ideals, ethical vices, as well as people's character and appearance.

The article by Piret Voolaid, *Children's Funny Remarks in the Field of Linguistic Humour Theory*, analyses children's spontaneous sayings, recorded during daily activities and interaction, as well as the answers given to the teacher's questions, guided by the latter's interest, and proposes the theoretical mechanisms of humour they are based on.

The article titled *Giants in Transmedia* by Mare Kõiva and Andres Kuperjanov discusses transmedia narratives based on giant lore, which is described using examples from folkloristics and transmedia dissemination. Giant lore, particularly the national epic *Kalevipoeg*, a core text of Estonian culture, has generated numerous transmedially circulating texts and various contemporary forms. The article provides a closer look into spontaneous transmedia narratives.

Yuri Berezkin and Evgeny Duvakin in their article *The Captive Khan and the Clever Daughter-in-Law* analyse in detail the tale types *The Encoded Message* and *The Big Bull*. Both find correspondences in Balto-Finnic traditions, which allows us to discuss them in the context of previously unrecognised or poorly studied parallels between the Caucasus and Northern Europe. The approximate date of their links is the second part of the first millennium A.D.

The contribution by Ekaterina Velmezova, titled *The Estonian Language as Presented in the* Great Soviet Encyclopaedia*: Language Descriptions and Ideology*, compares the articles about the Estonian language published in 1934, 1957, and 1978. The conclusion is that the article in the first edition of the encyclopaedia is the most ideologically marked and reflects the influence of Marrism on its authors, whereas the language is described to a lesser degree.

Editors

http://dx.doi.org/10.7592/FEJF2016.64.jumal

ESTONIAN *JUMAL* 'GOD'

Anni Jürine, Karl Pajusalu, Renate Pajusalu, Ilona Tragel, Ann Veismann

Abstract: The article focuses on semantics and the use of the word *jumal* 'god' in Estonian. At first, we give an overview of the etymology of *jumal* 'god' and its meanings in Standard Estonian and Estonian dialects. Then we examine *jumal* 'god' as a linguistic unit associated with different constructions and collocational patterns, bleaching of the meaning, and usage in fulfilling different communication goals. In our study we show how *jumal* 'god' can refer literally to the higher power of Christianity or some other religion but it is also used in the weakened, bleached sense, meaning 'very, completely' or 'total, absolute', and in combination with certain verbs, meaning 'random', 'indefinite'.

Keywords: collocations, discourse analysis, Estonian, etymology, folk linguistics, semantics

INTRODUCTION

This article is inspired by Urmas Sutrop's long-time linguistic and ethnological interest in the concept *god* (see for example Sutrop 2003, 2004a). We endeavor to give a comprehensive overview of the etymology of the Estonian word *jumal* 'god' and its use in modern Estonian and Estonian dialects. We therefore examine *jumal* 'god' not as a theological concept but as a linguistic unit associated with different constructions and collocational patterns, bleaching of the meaning, and usage in fulfilling different communication goals.

In modern Estonian, *jumal* 'god' can refer literally to the higher power of Christianity or some other religion but it is also used in the weakened, bleached sense 'very, completely' or 'total, absolute'. A metaphoric meaning 'object of respect or desire', for example, *Kunst jäi tema jumalaks* 'Art remained his/her god' (EKSS 2009), can also be found. Since there is no grammatical gender in Estonian it is not possible to determine based on the linguistic material whether *jumal* 'god' is male or female.

To show all of the variation in the use of the word *jumal* 'god' we have gathered data from different corpora of Estonian (etTenTen, University of Tartu Corpus of Written Estonian, University of Tartu Corpus of Spoken Estonian, University

of Tartu Corpus of Estonian Dialects). First, we will present the etymology of the word *jumal* 'god' and its distribution in the Finno-Ugric languages and dialects of Estonian, then we will examine the verb collocations of *jumal* 'god', i.e., answer the question *what does god do*? In the second part of the article we will concentrate on the uses of the word *jumal* 'god' where it does not literally mean a higher religious power but has aquired new functions in exclamatory sentences as an intensifier of other nouns, and in the constructions *jumala eest* and *jumala pärast* (both could be translated as 'for God's sake').

ETYMOLOGY AND DIALECTS

The word *jumal* 'god' is known in all Finnic languages and the word *jumo* is used actively in this sense in Mari languages (EES 2012: 100–101). In Permic, Ugric and Samoyed languages the word is not attested, showing that it does not belong to the oldest layer of vocabulary of the Finno-Ugric languages. Nevertheless, some sources have considered it a native word (SSA I 1992: 247). More likely, it is an Indo-Iranian loanword. Jorma Koivulehto has proposed an original stem **diyumna*, cognate of Sanskrit *dyumān-* 'heavenly, shining, radiant' (Koivulehto 1999: 228; see also Uibo 2014: 190–191). Recently, the origin of the word *jumal* 'god' has been examined in detail by Ene Vainik (2014) who also considers the (Proto)Indo-Iranian origin likely but deems the most suitable original stem to be **jumV*, meaning 'twin'. In addition, referring to Karulis (1992: 361), Vainik compares the word with Sanskrit derviation *yamala* 'in pairs, doubly' and Prakrit *yamala-* 'twins', but also Latvian *jumala* 'abnormally fat female (also when expecting twins)' (Vainik 2014: 21–26). However, Vainik is not completely convinced whether the word was borrowed into the Finnic languages in the form **jumala* and additionally presents the options that the word could originally have been a derivation with a locative suffix *-l(a)* or a compound word with a locative meaning (**jum+ala* 'God's place') (ibid.: 25). In all Finnic languages the word refers to a supernatural being, i.e., personified god, probably being borrowed as such already into the Proto-Finnic (or an earlier form of it).

In the development of the different meanings and usage connections of the word *jumal* 'god' the uniqueness of it as a mental category in the world-view of the language has played an important role (Sutrop 2003). It is evident that *jumal* 'god' has not been used in the Finnic languages as the name of a specific god (c.f. for example Taarapita, Sutrop 2001a), rather, it has been used as a generic term for a divine being, both in the abstract sense of a heavenly God as well as the concrete sense of a statue of God (for the latter, see Sutrop 2004a).

The word *jumal* 'god' is attested in all dialects of Estonian (EMS II, 6: 174–178), occurring in the forms *jumal ~ jummal ~ jummal´ ~ d´ummal´~ jomal ~ jommal ~ jõmmal´* etc. The several different dialectal developments of the word's phonetic form may be related to its expressive content, for example the gemination of the stem consonant *-m-* and the palatalization of the final consonant *-l* (*jummal´ ~ jommal´~ jõmmal´*). In Estonian dialects the word refers to the Christian God, an omnipotent supernatural being, a pagan deity, a faery, figuratively also nature, destiny, an omniscient person, also a statue of god and an icon. In addition to being used as the main word of the phrase it can be widely used in dialects as a modifier signifying something ordained by God, natural, inevitable, independent of man's actions, e.g., Kod *suri jumala surma, ega tädä es tapeta* 'died of natural causes, he/she was not killed, lit. died God's death', Nõo *üte jumala päevä kõik* 'God's days all (i.e., all days are the same)', Rõu *kiä tiid kas um tä jumala rist vai tulõ esi'* 'who knows if this/he/she is God's cross (i.e., destined to be) or comes by itself (spontaneously)'. The expression *jumala ilm* 'God's weather' may refer to both good weather and bad weather, storm, thunder: IisR *ninda ilus pühäbäne jumala ilm* 'such nice Sunday-like God's weather', Mus *vali jumala ilm oli väljas* 'there was stormy God's weather outside'.

In dialects the word *jumal* 'god' also occurs in many fixed expressions and exclamations, especially expressing startlement and regret: Mar *armoline jumal* 'merciful God', Lüg *jumal paraku* 'God make better', Kaa *jummal´ oitka* 'God keep', Mar *tule sa jumal appi* 'God help', Lüg *kie jumal sedä tiab* 'who the God knows that', etc. The expression *jumala eest* (lit. from the front of God) means 'certainly, definitely, indeed': Krk *jumala iist sii om õigus* 'indeed this is true'; in a negative sentence its meaning is 'by no means, under no circumstances': Jäm *jumala eest ma pole teind seda* 'by God, I have not done this'; the expression *jumala pärast* 'for God's sake' occurs in confirmations, but often also in warnings: Krk *jumale pärast, ära sa mitti tetä usu* 'for God's sake, do not belive him/her'. In requests also the expression *annaks ~ aitaks jumal* 'God give ~ help' is used: Pee *annaks jumal vihmukest, et saaks selja sirutist* '(colloquial) God give rain so we could straighten our backs'.

Metaphoric use of the word *jumal* 'god' with different verbs is diverse. In the South Estonian dialects *jumaladõ ~ jumalahe ~ jumalilõ minema* 'go into God' refers to sunset: Har *päiv nakas jumaladõ minemä* 'the sun started to set'; *jumalan ~ jumalah olõma* 'be in God' means 'to have set': Har *päiv om jumalan* 'the sun has set'. This metaphor is not known in other European languages, however, it can be connected with the Estonian idiom *päike läheb looja* 'the sun is setting' (the ambiguous word *looja* means 'the Creator', too, see EES 2012: 249). However, in European languages the idiom *jumala karja ~ kohtu ette*

minema 'go to God's flock ~ to the front of God's justice' has a lot of equivalents. Such phrases refer to dying in both North and South Estonia: Rei *läks jumala karja* 'went to God's flock', i.e., 'died', Krk *sii om jumale karja ärä lännu joh* 'he/she has already gone to God's flock', i.e. 'he/she has already died'; *jumala karjas ~ kohtu ees olema* 'be in God's flock ~ in front of God's justice' is to be dead: Lai *see ammu jumala karjas* 'that one has been in God's flock for a long time', i.e. 'that one died a long time ago'.

A widespread greeting, especially to one working in the field, was *jumal appi* 'God to help' to which one always replied *aita jumal* 'help, God': Kod *kui tuleb väljäle tüü juure, ütleb: jumal appi; tõene ütleb: aita jumal ~ jumal tarvis* 'when (a person) comes to the field where work is being done he says: *jumal appi* 'God to help'; the other replies *aita jumal* 'help, God' or *jumal tarvis* 'God needed".

As an adverb of emphasis, *jumala* – the genitive form of the word *jumal* – is used (EMS II, 6: 178). It usually expresses the meaning 'completely, fairly, quite': Khk *see oo nii jumala õige mis ma räägi* 'it is so God's true what I'm saying', i.e., 'it is completely true what I'm saying', *Se tu oll´ jumala õnn, et ma är pässe* 'it was God's luck that I escaped', i.e., 'it was quite lucky that I escaped'. In the case of negation, it intensifies the meaning 'not at all': Kse *ma ei tea sest jumala põrmu* 'I do not know God's ashes of this', i.e., 'I do not know anything at all of this'. As an adverbial of time accompanied by the words *iga* 'every' or *kõige* 'most', the word *jumala* 'God's' expresses the lengh of a period of time: Har *nuidõgõ käve kõgõ jumalõ talvõ* 'with these I walked the whole God's winter', i.e., 'I wore these the whole winter'.

Some other case forms of the word *jumal* 'god', its derivations and phrases containing it that have merged into compounds and shortened have also acquired independent meanings (EMS II, 6: 178, 182). For example, the comitative form *jumalaga* 'with God' means 'good-bye': Pöi *jäta vanaisa jumalaga koa* 'leave grandfather with God, too' i.e., 'say good-bye to grandfather, too'; deminutive derivation *jumaluke(ne)* or its partitive form *jumalukest* is used as an interjection: Kod *küll one kua irmus surm, jumalukene* 'it is indeed a horrible death, dear God', Khk *jumalukest, mis sa sest ka saad, seda oo ju nii vähe* 'dear God, what will you get from this, there's so little of it'; the interjection *jumalime* (~ *jumala ime* 'God's miracle') ~ *jumalmine* ~ *jumaline* ~ *jumalimel* (< **jumala nimel* 'in God's name') developed via coarticulative abbreviation: Rei *tere jumalimel* 'hello, God's miracle (surprised)'.

No clear difference can be observed between dialects in the use of the word *jumal* 'god' in different fixed expressions and its lexicalization in new figurative meanings. However, exceptionally numerous and varied examples originate in the Southern and Insular dialects as well as South Estonian linguistic enclaves in Latvia.

WHAT DOES GOD DO (AND WHAT DOES HE NOT DO)?

In this section, we set out to answer the question what does *jumal* 'god' do on the Estonian Internet? In other words, we search for verbal collocates of *jumal* 'god'. The following analysis is based on the data extracted from the etTenTen corpus, the currently largest corpus of Estonian, compiled of 270,000,000 words of texts collected from Estonian web sites[1]. As we are interested in god's activities, we searched for clauses that include the singular nominative form of the word *jumal* 'god' as a noun and a proper noun and a verb in the 3rd person singular form. Because of reasons of time and space, the query was restricted to verbs in the indicative mood, present tense, and affirmative polarity. The verb *olema* 'be' was excluded by search parameters.

Table 1. *Top 20 verbal collocates of* jumal *'god' based on absolute frequency of co-occurrence.*

#	Verb lemma	Meaning	Absolute frequency
1	*teadma*	know	612
2	*andma*	give	534
3	*tahtma*	want	460
4	*tegema*	do	290
5	*saama*	get	276
6	*armastama*	love	249
7	*ütlema*	say	224
8	*võima*	can	214
9	*kutsuma*	call	181
10	*aitama*	help	147
11	*nägema*	see	144
12	*rääkima*	speak	143
13	*tulema*	come	137
14	*ootama*	wait	111
15	*lubama*	allow	109
16	*juhtima*	direct	106
17	*pidama*	have to	106
18	*võtma*	take	104
19	*juhatama*	guide	97
20	*soovima*	wish	97

Table 2. *Top 20 verbal collocates of* jumal *'god' based on the log-likelihood measure.*

#	Verb lemma	Meaning	Log-likelihood score
1	*armastama*	love	1,216
2	*teadma*	know	1,171
3	*tahtma*	want	845
4	*saama*	get	683
5	*andma*	give	654
6	*juhatama*	guide	502
7	*kutsuma*	call	472
8	*tegema*	do	376
9	*ilmutama*	reveal	354
10	*pidama*	have to	350
11	*tulema*	come	348
12	*õnnistama*	bless	340
13	*võima*	can	329
14	*juhtima*	direct	249
15	*päästma*	save	230
16	*võtma*	take	206
17	*karistama*	punish	203
18	*hoolitsema*	care (for)	200
19	*kõnetama*	bespeak	186
20	*vastama*	answer	156

The search resulted in 8277 examples, some of which included two or more clauses that contained *jumal* 'god' (as the subject) and a finite verb. After cleaning the data, we ended up with the final dataset of 8329 examples. Within these 8329 examples *jumal* 'god' co-occurred with 598 verb lemmas. In the following, we present the top verb collocates of *jumal* 'god' based on absolute frequency of co-occurrence (Table 1) and top collocates based on collocational strength (calculated with the log-likelihood measure) (Table 2).

The tables suggest that the two methods of determining collocates give partly overlapping results. The verbs that rendered the strongest collocates based on absolute frequency as well as collocational strength are given in shaded cells. It can be observed that on the Estonian Internet, *jumal* 'god' most commonly *armastab* 'loves', *tahab* 'wants', *teeb* 'does', *teab* 'knows', *annab* 'gives', *võib* 'can', *kutsub* 'calls', *tuleb* 'comes', *juhib* 'directs', *juhatab* 'guides', *peab* 'has to' and *võtab* 'takes'.

The data suggests that the verbs that are frequent collocates of *jumal* 'god' based on the absolute frequency of co-occurrence (see Table 1), but have weaker collocational strength, tend to be more frequent in general and express neutral activities (*ütlema* 'say', *aitama* 'help', *nägema* 'see', *rääkima* 'speak', *ootama* 'wait', *lubama* 'allow', *soovima* 'wish'). However, some of those verbs, i.e. *nägema* 'see' (as in example (1)) and *aitama* 'help' (as in example (2)), are associated with specific features of God, i.e. the omnipresence and omnipotence.

(1) *Jumal* ***näeb*** *kõike.*
'God sees everything.'

(2) *"Ma teadsin, et Jumal* ***aitab*** *meid sealt välja," kirjeldas oma läbielamisi kolmas päästetud kaevur 40-aastane Mario Sepulveda.*
'"I knew that God would help us get out," said 40-year old miner Mario Sepuvelda when describing what he had been through.'

On the other hand, not all of the strong collocates are necessarily among the most *frequent* collocates of *jumal* 'god'. Table 2 also includes verbs that are not that frequent based on their absolute co-occurrence with *jumal* 'god'. These verbs have strong associational strength with *jumal* 'god' due to their low frequency in general. Such verbs include lemmas that are rather specific to the agent at hand, e.g. *ilmutama* 'reveal' (see example (3)), õnnistama 'bless' (see example (4)), *päästma* 'save', *karistama* 'punish', *hoolitsema* 'care (for)'.

(3) *Jumal* ***ilmutab*** *end nõrkades, allasurututes, silmapaistmatutes.*
'God reveals himself in the weak, oppressed, mediocre.'

(4) *Jumal* ***õnnistab*** *meid iga päev rikkalikult.*
'Every day God blesses us abundantly.'

When the common activities of *jumal* 'god' are compared to that of humans we see little overlap. The list of common activities of humans is based on Veismann et al. (2002) who used the list task and cognitive salience index (Sutrop 2001b) to determine the most salient Estonian verbs. The comparison of the top 20 rankings in the two studies shows that only three verbs overlap – *tegema* 'do', *tulema* 'come', and *armastama* 'love' (Veismann et al. 2002: 320). The small amount of shared verb lemmas may be due to the fact that the results have been obtained using different methods applied to achieve different goals. The list task in Veismann et al. (2002) was employed to determine the verbs that belong to the core vocabulary of Estonian, whereas the present corpus study attempts to investigate activities of a specific (inhumane) agent. Therefore, it is not unexpected that the (human) subjects produced verbs that refer to more mundane activities, such as eating, sleeping, drinking (Veismann et al. 2002: 321) but *jumal* 'god' is mainly associated with more abstract activities (e.g. guiding, knowing, wanting). In the following sections, the main activities of *jumal* 'god' are discussed in more detail.

God loves us ... no matter what?

Armastama 'love' is the most likely verb to co-occur with *jumal* 'god' according to the log-likelihood measure (see Table 2 above). Based on its absolute frequency of co-occurrence, it holds the 6th position with 249 instances (see Table 1). Most of the examples of *armastama* 'love' in our dataset pertain to the Christian context, referring to the all-loving God (as exemplified in (5)). However, there are also examples where conditions are listed, under which god would love someone or something (see example (6)).

(5) *Pidada meeles, et kõiki inimesi, sh kristlasi,* ***armastab*** *Jumal tingimusteta ning sõltumata seksuaalsest orientatsioonist ja sooidentiteedist.*
'One must keep in mind that God loves all people, including Christians, unconditionally and regardless of their sexual orientation and gender identity.'

(6) *Sõnakuulelikku* ***armastab*** *jumal, sõnakuulmatut ta vihkab.*
'God loves the obedient, but hates the disobedient.'

God knows and guides

Teadma 'know' is the most frequent verb (612 occurrences) and one of the strongest collocates of *jumal* 'god'. The data includes cases where *teadma* 'know' is used in its literal meaning (as in (7)) where the combination of the verb and *jumal* is used to express the ability of God to know what somebody is doing. However, in addition, the string *jumal teab* 'god knows' is also used as a lexicalized whole and it expresses indefiniteness. For instance, in (8) *jumal teab* 'god knows' means that nobody knows what the social workers are up to. The semantic shift of the expression assumedly takes place in contexts observable in (9). The shift is based on the assumption that god (only) knows [something] and that we, i.e. humans do not. Therefore, nobody knows.[2] However, this type of change is not confined to *jumal* 'god' but can be found in case of other 'taboo agents' (e.g. *kurat* 'devil') as well (see Kehayov 2009).

(7) *Tõesti, Jumal* ***teab*** *täpselt, mida te teete.*
'Really, god knows exactly what do you do.'

(8) *Sotsiaaltöötajad tegelevad jumal* ***teab*** *millega.*
'Social workers are up to god knows what.'

(9) *Palju selle lammutatud süsteemi taasloomine jälle maksma läheb, Jumal* ***teab****, Pille Saar mitte.*
'How much does it cost to rebuild the demolished system? God knows, Pille Saar doesn't.'

A similar semantic shift has taken place with the verb *juhatama* 'guide'. Most commonly, it appears in the examples such as (10) where the string (*nagu*) *jumal juhatab* 'as god guides' has developed a holistic meaning 'randomly'. Thus, in this case too, the semantic shift is based on the interpretation that the manner in which god guides things is unbeknownst to us, humans. However, in some rare cases, the verb *juhatama* 'guide' also appears in its concrete meaning (as in (11)). Such cases resemble examples with another verb stemming from the same root – *juhtima* 'direct' (as in (12)).

(10) *Tõsi, ühe kaupluse juures küll vahetati pudelid raha vastu, kuid seda nii nagu jumal* ***juhatab*** *ehk täiesti ettearvamatutel aegadel.*
'True, at one of the stores, we managed to exchange the empty bottles for money; but this happens at totally random times.'

(11) *Katolikus ühiskonnas puudub üksikisiku vastutus, sest Jumal* ***juhatab*** *nagu peab.*
'In catholic society, an individual has no responsibility because God guides [people] in their actions.'

(12) *Inimene mõtleb, jumal* ***juhib****.*
'Man thinks, God directs.'

God often gives, but less often takes

Andma 'give' is the second most frequent verb in our dataset. The combination of *jumal* 'god' and *annab* (give-3SG) has 534 occurrences in our data, and the verb holds the 5th position as for the collocational strength. The entities given by god include various types, e.g. God gives people peace (as in (13)) or an overview (as in (14)). Majority of the instances of the phrase *jumal annab* 'god gives' are related to religious contexts (as in (13)).

(13) *Jumal* ***annab*** *rahu südamesse ja hinge.*
'God gives [us] piece of mind.'

(14) *Jumal* ***annab*** *siin ülevaate universumi elanike genealoogilisest päritolust.*
'Here god gives an overview of the genealogical heritage of the habitants of the universe.'

The opposite action, taking, is however more infrequent in god's actions: the instances of the phrase *jumal võtab* 'god takes' are about 5 times less frequent, altogether 104 instances. Despite of this, *võtma* 'take' is still considered to belong to the group of top collocates of *jumal* 'god'. However, it should be noted that the verb is mostly not used in the literal meaning but occurs as a part of the phrasal verb *võtab vastu* 'welcomes, accepts', as in (15) or *võtma* 'take' + V_{INF1} construction *võtab kuulda* 'take heed', as in (16) (cf. also Tragel et al. 2015).

(15) *Jumal* ***võtab*** *KÕIK* ***vastu****, lollid kaasaarvatud!*
'God accepst everybody, even fools.'

(16) *Et isegi kui üks inimene palvetab,* ***võtab*** *Jumal seda* ***kuulda****.*
'Even if one individual prays, God will hear them.'

God wants

Tahtma 'want' is the third most frequent verb in our dataset (460 occurrences) and holds the 3rd position in the ranking based on collocational strength. In Estonian, the verb *tahtma* 'want' has many meanings and is grammaticalized as far as to the postmodal avertive function (see Habicht et al. 2010). Wanting in general is an intrasubjective concept expressing a desire and/or volition. According to our data, the desires of god resemble that of humans (see examples (17) and (18)).

(17) *Jumal* ***tahab*** *teha selliseid asju, mis haaravad inimeste tähelepanu.*
'God wants to do the kind of things that attract people's attention.'

(18) … *Jumal* ***tahab****, et minust saaks president.*
'God wants me to become a president.'

God comes, but does not go

The Estonian core verbs (see Tragel 2001) are rather well presented among the top collocates of *jumal* 'god'. For instance, the top collocates include the following core verbs – *tegema* 'do', *tulema* 'come', *saama* 'get', *andma* 'give', *võtma* 'take', *pidama* 'have to', *võima* 'can'. It is quite expected that the verbs that are very general, grammatical, and frequent in Estonian are also present in the context of god's activities. After all, god made man after his own image, or rather, the other way around. In each case, it is not surprising that their core activities overlap. Nevertheless, some core verbs, especially verbs of movement do not express activities common to *jumal* 'god'. The data suggest that the combinations of *jumal* 'god' and *läheb* (go-3SG) and *jumal* 'god' and *käib* (go-3SG) only appear 17 and 11 times respectively (see example (19)). Thus, it seems that *jumal* 'god' is fonder of actions than motion. However, *tulema* 'come', which occurs on 137 occasions and also belongs to the strongest collocates is an exception. *Tulema* 'come' typically appears in the examples where *jumal* 'god' moves (in a physical or abstract sense) towards humans (see (20) and (21)).

(19) *Tean nimelt, et Jumal* ***käib*** *mu ees ja kõrval.*
'In fact, I know that God walks in front of me and beside me.'

(20) *Jumal* ***tuleb*** *maa peale!*
'God comes to Earth.'

(21) *Jumal ise* ***tuleb*** *Jeesuses Kristuses meie juurde ning seda mitte vaid kord, möödanikus, vaid pidevalt.*
'God himself comes to us in the body of Christ; and not only once in the past, but constantly.'

In sum, the data suggests that in addition to the more common, neutral verbs, the collocates of *jumal* 'god' also include verbs that are more agent (god) specific (e.g. *päästma* 'save', *ilmutama* 'reveal', õnnistama 'bless', *armastama* 'love', and *karistama* 'punish'). These verbs reflect the dominantly Christian nature of *jumal* 'god' on the Estonian Internet. The data also suggests that *jumal* 'god' is more fond of action than motion and that the top verb collocates are quite abstract as for the actions they refer to. As god walks in mysterious ways, the combinations of *jumal* 'god' and certain verbs (*teadma* 'know', *juhatama* 'guide') have acquired holistic interpretations, meaning 'indefinite' or 'random'.

JUMAL IS NOT ALWAYS 'GOD'

Although in modern language use when using the word *jumal* 'god' the speaker is not always referring to god as a higher power Estonians' language use is nevertheless based on a Christian background (see also Riistan 2006). However, as in other European languages the use of God's name differs from that of earlier times when even naming the higher power could have great consequences. One of the speech acts associated with *god* is swearing. People "swear by some higher force or somebody; we swear that something is so; we swear to do something; we swear at something or somebody; and we swear simply out of anger, disappointment, or frustration" (Hughes 2006). Swearing, originally having ritual meaning, has developed into emotional exclamatory sentences and intensifiers of scalar qualities.

Jumal 'god' as an interjection

The noun *jumal* 'god' can occur as an emotional interjection comprising an independent utterance (mainly in the form *oi jumal!* 'oh God!'; *issand jumal!* 'Lord God!', *jumal küll!* 'God indeed!' or *jumaluke!* 'dear God!') or belonging to a larger utterance as a syntactically free adjunct. In this case the noun *jumal* 'god' is desemanticized, the utterance lacks internal structure (c.f. *Taevake!* 'Heavens!', *Oh sa tuline tuhat!* lit. 'Oh you fiery thousand!', *Kurat!* 'Devil!', *Sa armas aeg küll!* lit. 'You dear time!' etc, see EKG II 1993: 182). Similar constructions in dialects were discussed above.

The exclamative utterances containing *jumal* 'god' mostly express some type of emotion and are placed within the syntactic unit that specifies the object of that emotion. These exclamatory sentences originate from an appeal to some higher power who should then help the speaker with his/her predicament related to the emotion. Sometimes the construction also explicitly contains a term of address *sa* 'you' (*oh sa jumal!* 'oh you God!', *oh sa püha jumal!* 'oh you holy God!' etc.). Historically, such address has been deontic (the speaker tried to influence the higher power to do something), in the modern secularized world it is rather the expression of the speaker's subjective emotion (Arnovick 2000: 4).

While it is probably impossible to list all the emotions that the phrases containing *jumal* 'god' represent, a few types can be observed in the data. For example, one distinct group is comprised by situations of opposition where the speaker utters a sentence beginning with an exclamatory phrase containing *jumal* 'god' that opposes the partner's previous statement (22–24). The phrase containing *jumal* 'god' can also precede a reproach (25).

(22) *R: mul akkab juba alb kui ma seda suurt söömist $ vaatan $.*
'I get sick already when I look at all this eating.'

E: ***o:i jumal*** *(.) mul läheb aina paremaks. (1.0) [S]*
'Oh God, I feel better and better.'

(23) ***jumal küll*** *ega ma sis nüüd ei ole (-) mingi siuke hehe $ röövlitädi [S]*
'For God's sake, I'm not some such robber-auntie.'

(24) ***ah jumalukene*** *ega ma tervet lugu soa meelde jätta sis. [S]*
'Ah dear God, I can't remember the whole story, can I.'

(25) *Issand* ***jumal****! Sa ikka alles minemata!*
'God! You still haven't gone!'

An exclamatory phrase containing *jumal* 'god' can also occur at the end of the sentence. In example (26) it seems to express mainly irritation and occurs after the partner has initiated self-repair and coveys displeasure that the repair was needed at all.

(26) *M: $ ma enam ei tee. $ (laughter in the background)*
'I won't do it again.'

T: $ autor krooksub. $ hehe
'The author belches.'

M: $ mis autor ma olen. $
'I'm no kind of author.'

T: nojah, [see kes LITEREERIB, ***jumal küll****.] [S]*
'Alright, the one who transcribes, for God's sake.'

Since exclamatory phrases containing *jumal* 'god' are widespread and stylistically fairly neutral in everyday speech they are often used for marking the occasions when somebody else's words or thoughts are being reported. In example (27) the speaker repeats Riho's words which, according to the speaker, Riho had said in a surprised manner. Characteristically to spoken language the reported speech starts with the conjunction *et* 'that' but since the intonation changes and the pronoun *teil* 'you' probably refers to Riho's conversation partner the illusion of direct speech is created for the listener. *issand jumal* lit. 'Lord God' helps to support this interpretation.

(27) *T: .jajah? (.) siis lähed öösel [uut otsima.]*
'Yes, then you go at night to look for a new one.'

L: [ei saa peksa.] (.) Riho imestas=et (1.0) @
issand jumal *teil need veel alles. @ hehe $ ei ole ikka katki läind. $ (1.2) [S]*
'Won't get beaten up. Riho wondered that God you still have these, they still haven't broken.'

Sometimes, thoughts that may never have been uttered can be presented as reported speech. The speaker conveys the event as if it took place at the time of speech and the phrase containing *jumal* 'god' adds emotional authenticity. In example (28) the parts *oh jumal* 'oh God' and *mis nüüd saab siis* 'what will happen now then', starting with the conjunction *et* 'that', specify the content of the phobia that is mentioned earlier, *oh jumal* 'oh God' indicates that the phobia is emotionally loaded. Example (29) does not contain a conjunction and the strongest device creating the appearance of direct speech is the phrase *issand jumal* 'lit. Lord God'. In examples (27), (28) and (29) it can be seen that the phrase containing *jumal* 'god' combines with deictics (*teil* 'you' and *need* 'these', *nüüd* 'now', *see* 'it' and *mina* 'me', accordingly) and the verbs in the sentences are also in present tense (although the rest of the context, characteristically to a narrative, is in past tense) which also supports the impression of direct speech.

(28) *nojaa aga aga asi on selles, et kui nüd eksole sattuda niuksesse ee foobiasse, eksole=et* ***oh jumal*** *et mis nüüd saab sis*
'Yes, but the thing is that if you get into such a fobia, right, that, oh God, what will happen now then'

(29) *A: [jajaa] (.) aa, mina vaatasin enne ülikooli ees=et, kellele (.) kellele=se=Jaak nii viipab onju, nii sõbralikult=et ma olin (.) ma olin ka nii kaugel, [s=lähedal] vaatsin* ***issand jumal*** *mis [see on,*
'Yes, yes, but I was looking earlier in front of the university that who does that Jaak wave to, right, in such a friendly way that, I was, I was also so far, near, I looked that, God, what is it,'

*M: [(-)] [****issand*** *kas see tõesti olen mina (-)]*
'Lord, is it really me'

Such phrases containing *jumal* 'god' that intensify the emotion being described also occur in written text that attempts to convey the emotionality of spoken language. Example (30) refers to the speaker's feelings which are in fact only characterized by the phrase *issand jumal* lit. 'Lord God'.

(30) *Kui ma fentanüüli proovisin, siis ma tundsin, et* ***issand jumal****, kas on tõesti võimalik, et kõik halvad emotsioonid on kadunud.*
'When I tried fentanyl I felt that, God, is it really possible that all the bad emotions have dissappeared.'

Jumal **'god' as an intensifier**

The genitive form of the word *jumal* 'god' has grammaticalized into an intensifying adverb meaning 'very' and modifying an adjective (see also EKSS 2009). It was mentioned above that this use is also present in dialects. *Jumala* 'god's' can probably intensify any adjective but it is primarily found with scalar[3] adjectives (31–34).

(31) *Siis nägin Markot ja ajasime juttu ja* ***jumala*** *kihvti juttu võib see tüüp ikka ajada. [I]*
'Then I met Marko and we talked and that guy can really talk.' Lit. 'can talk God's amazing talk'

(32) *tal on* ***jumala*** *sirged siuksed panni juuksed on muidu onju [S]*
'He/she has totally straight hair like a pan otherwise, right.' Lit. 'God's straight hair'

(33) *ja=ja siis Summing seal noh minuga* ***jumala*** *suur < sõber > onju [S]*
'And-and then Summing there, well, really great friend with me, right.' Lit. 'God's great friend'

(34) *Anni juurde seal on **jumala** ilus onju [S]*
'To Ann's place it is really beautiful there right.' Lit. 'God's beautiful'

Jumala 'god's' can also intensify scalar adverbs (35).

(35) *Aga pooled neist või vähemalt paljud neist olid kuidagi tiidakil ja Klaara maja "paraad" -uks oli märgatavalt tiidakil, kaks paekivist trepiastet servast murenenud ja ülemine **jumala** lõhki, …*
'But half of them or at least many of them were somehow askew and the front door of Klaara's house was noticeably askew, two limestone steps crumbled at the edges and the upper one completely split, ...' Lit. 'God's split'

Jumala 'god's' can occasionally also occur as a modifier preceding a noun, thus functioning as an adjective meaning approximately 'true, real, great'. However, this use is considerably less frequent than its use as an adverb. Pajusalu (2006) only found one noun – õnn 'luck' – with *jumala* 'god's' as its modifier (36) in the spoken language data whereas *kuradi*, the genitive form of the word *kurat* 'devil' which has similar functions at first glance typically occurred as a modifier (*kuradi paarike* 'damn couple', lit. 'devil's couple', etc.)

(36) *aga **jumala** õnn=et ´Kadaka ´Tiia see (.) mees mehel tulid brigaad tuli ´krohvima ga´raazhi minu kõrval=ja [S]*
'But it's God's luck that Kadaka Tiia's that husband had a team come to plaster the garage next to me and'

Jumala 'god's' as a modifier is more frequent in written language but mostly preceding abstract nouns that can be interpreted as scalar and that could also be preceded by adjectives *suur* 'big' or *täielik* 'total'. For example, there are occurrences of the noun phrase *jumala tõsi* 'God's truth' (37).

(37) *Hoolimata spontaansest teietusest oli see **jumala** tõsi ning ei johtunud kindlasti mitte joodud pudelist õllest, see oli isegi minu jaoks tühine kogus.*
'Despite the spontaneous use of *teie* [polite form of address] it was God's truth and was definitely not caused by the bottle of beer I had drunk, that was a paltry amount even for me.'

There are also some sentences in the corpus where *jumala* 'god's' modifies a verb (38) although these are probably not acceptable as grammatical for all speakers of Estonian. Here as well the meaning is primarily intensifying.

(38) *Vaatasin enda kätt ja siis üks hetk **jumala** kangestusin.*
'I looked at my hand and then one moment I totally froze.' Lit. 'God's froze'

Phrases containing *jumala* 'god's' can also function as intensifiers. All the exclamatory phrases described above intensify the emotions being expressed. Sometimes such exclamatory phrases can also be used as modifiers following a noun (*sihuke naistemees et jumal hoia* 'such a ladies' man that God keep', 39).

(39) *Vahest on säherdune vennike, kui koduse lõa otsast pääseb, sihuke naistemees et* ***jumal*** *hoia! [I]*
'Sometimes such a guy, when he escapes from the leash at home, is such a ladies' man that God keep!'

As in dialects, in written language there is also such use where the meaning of *jumala* 'god's' is rather 'natural, all-encompassing'. Accompanied by time units *jumala* 'god's' occurs in the construction *iga jumala päev / nädal / ?* 'every God's day/week/?' (40, 41) where the function of *jumala* 'god's' is to intensify the meaning of the quantor *iga* 'every'.

(40) *Ma ju tunnen seda nagu oma kodu, iga* ***jumala*** *päeva veedan ma seal, olgu külm või sadagu vihma.*
'I know it like my own home, every God's day I spend there, be it cold or raining.'

(41) *Iga* ***jumala*** *nädal viis ta mind, isa tähendab, ratsutama.*
'Every God's week he took me, father, that is, riding.'

Idiomatic phrases *jumala eest* and *jumala pärast*

As already mentioned above the word *jumal* 'god' occurs in many fixed phrases, two very frequent of which are *jumala eest* and *jumala pärast* that could both be translated as 'for God's sake'.

It is a relatively rare phenomenon that words that are considered to be semantic antonyms can occur as synonyms in certain contexts. However, some examples can be found among highly polysemous words (e.g. *great* and *fine*). Estonian temporal adpositions *eest* 'ago, since' and *pärast* 'after' can be considered a semantically opposite pair (for example in sentences 42 and 43).

(42) *Kahe aasta eest* 'two years ago'

(43) *Kahe aasta pärast* 'after two years'

Both *eest* and *pärast* have grammaticalized through the concept of space. Both have become adpositions in the elative form of a noun (*ee-st* 'front-ELA'; *pära-st* 'back-ELA'). The etymological origin of the adposition *eest* is the noun **eδe*

(> *esi*, gen. *ee*) which means 'location in front, e.g. the strip or section of work ahead'. The word has cognates in several Finno-Ugric languages (for example Finnish, Veps, Ostyak, Vogul, Hungarian), also in the meaning of 'front part' or 'the area in front of the mover' (EES 2012). The etymological origin of the adpostion *pärast* is the noun **perä* (> *pära*, gen. *pära*) meaning 'location in back, rear part or end' (also with cognates in several Finno-Ugric languages, e.g. Livonian, Finnish, Komi etc.) (EES 2012).

The adposition *eest* has the (separative) meaning of spacial location (example 44, 45), the adposition *pärast* only expresses temporal and abstract relations.

(44) *Kohe* ***maja eest*** *läheb suur tee mööda.*
'A big road passes right in front of the house.'

(45) *Keegi ei tule mu* ***maja eest*** *lund ära koristama.*
'Nobody will come to clean up the snow from the front of my house.'

Both adpositions have numerous meanings/uses in the abstract domain, some of them being either similar (*eest* 'reward' and *pärast* 'cause, reason, because of') or (less frequently) the same (*võitlema millegi eest~pärast* 'fight for sth'). In addition, both adpositions *eest* and *pärast* can be accompanied by the noun *jumal* 'god' as a complement, thus forming synonymous fixed phrases *jumala eest* and *jumala pärast,* lit. 'for God's sake', an exclamation to emphasize a request. We will now go on to examine the use of these two phrases in written Estonian. The data was gathered from the Balanced Corpus of Written Estonian and the Estonian internet corpus etTenTen via *keeleveeb.ee*.

The query *jumala eest* yielded 95 sentences from the Balanced Corpus and 1060 sentences from etTenTen. Of these, 449 were analyzed (after randomizing the sequence of the sentences). The query *jumala pärast* yielded 38 and 390 sentences, accordingly, from the two corpora, leaving 351 sentences for analysis after the unsuitable sentences were removed. The sentences considered unsuitable were those where *jumala eest* or *jumala pärast* was either used in its literal sense (religious texts) or were exclamatory sentences containing only the one phrase. The religious sentences left out of the analysis were nevertheless briefly examined in order to get information about how the meaning of the phrase had developed. While the Spelling Dictionary of Estonian (ÕS 2013) also suggests that it is possible to spell *jumalapärast* as a single word, this was not included in the analysis since it would have only yielded six additional sentences, out of which only three would have been suitable for analysis. Grammatical polarity (negative, affirmative) and mood (indicative, conditional, imperative) of the verb in the clause associated with the phrase *jumala eest* or *jumala pärast* were analysed in the sentences.

As the data show, the phrase *jumala eest* is considerably more frequent in written Estonian than *jumala pärast*. Both *jumala eest* and *jumala pärast* are used to express various desires, commands and prohibitions.

The phrase *jumala eest* was most frequently used in combination with the negative of conditional mood (example 46), 117 times. Negative of imperative mood was used 96 times (example 47) and affirmative of indicative mood 100 times (example 48). *Jumala eest* was least frequent in combination with the affirmative of conditional mood (19 times, example 49).

(46) *Nägin, kuidas maja kõikus, ja mõtlesin: et ta* ***jumala eest*** *siiapoole ei kukuks.*
'I saw the building sway and thought: that it wouldn't fall this way for God's sake.'

(47) *Arva vähem ja* ***jumala eest****, ära rohkem kirjuta.*
'Have fewer opinions and for God's sake, don't write anything else.'

(48) ***Jumala eest*** *õige jutt ja mitu kuud tagasi oleks juba tehtud pidanud olema.*
'By God, that is true and it should have been done several months ago.'

(49) *Kui ma tohiksin nutta,* ***jumala eest****, ma nutaksin.*
'If I were allowed to cry, by God, I would cry.'

The phrase *jumala pärast* is most frequently accompanied by negation in imperative mood (example 50), 108 times, and negation in conditional mood (example 51), 87 times. *Jumala pärast* was least frequently accompanied by affirmative in conditional mood, 6 times.

(50) ***Jumala pärast*** *ära looda, et nad sind aitavad...*
'Don't, under any circumstances, hope that they will help you...'

(51) *Kõik valvavad kõiki, et keegi* ***jumala pärast*** *rohkem ei maksaks kui siin on heaks tavaks.*
'Everybody is watching everybody, so that no-one, under any circumstances, would pay more than is best practice here.'

Both phrases occur more frequently in negative sentences but *jumala pärast* favors negation more often (*jumala eest* 62% in the context of negation, *jumala pärast* 73% in the context of negation). *Jumala pärast* is frequent in the context of clearly verbalized prohibitions (*ärge jumala pärast tehke* 'do not, under any circumstances, do sth'), the use of *jumala eest* is characterized by prohibitions in conditional mood (*et nad jumala eest ei teeks* 'hope to God, that they wouldn't

do sth'). It is notable that *jumala eest* occurs relatively frequently also in the affirmative of indicative mood, where it occurs in the meaning of 'indeed' (example 48), whereas *jumala pärast* is rather rare in such context.

To sum up (see Figure 1), *jumala eest* occurs both in requests that something not be done (in conditional or imperative) and as a disjunct in the sense of 'indeed'.

Jumala pärast mainly occurs in imperative or conditional mood with requests that something would not be done/not happen. *Jumala eest* is more frequent and has semantically wider use than *jumala pärast*.

Looking at the sentences that were not included in the analysis and where the phrase occurred in a relgious context it can be observed that *jumala eest* is collocationally most frequently associated with conceiling and hiding (example 52) and escaping (example 53).

(52) *Aga sa ei saa* ***Jumala eest*** *midagi peita, mis mõte on siis isegi püüda seda teha?*
'But you cannot hide anything from God, then what is the point of even trying to do that?'

(53) *Ma põgenesin palju aastaid* ***Jumala eest***.
'For many years I was escaping from God.'

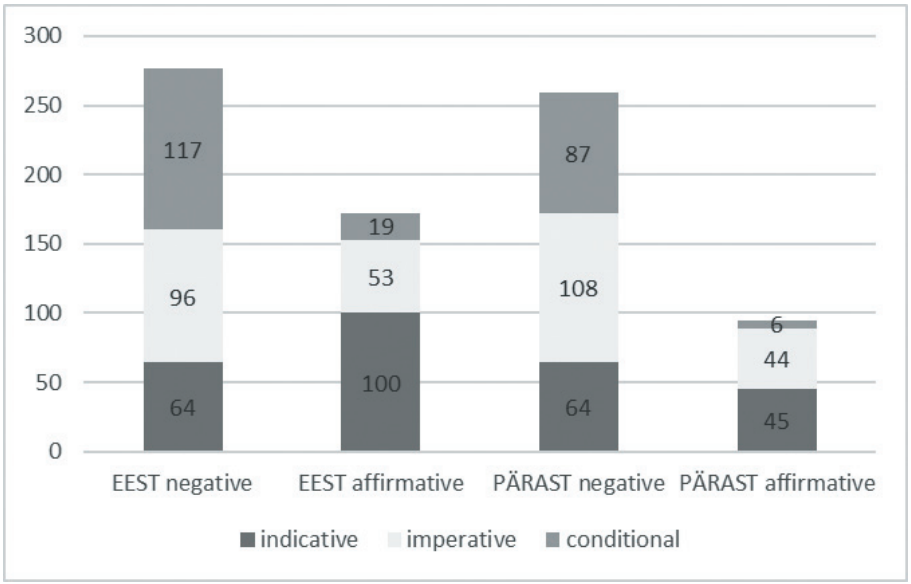

Figure 1. Jumala eest *and* jumala pärast *in Estonian Balanced Corpus and etTenTen.*

Jumala pärast mainly occurs in a context where something is being done because of God, God is the motive for doing something (for giving something up, fighting, loving etc.). It could be said that in the religious context the adpositions in the adpositional phrases *jumala eest* and *jumala pärast* have retained a certain connection with the original, spatial meaning: when escaping or hiding, the escaper moves in front of the one from whom he/she is trying to escape; while *pära* 'rear end' is the part of a vehicle that, when one applies force to it, can be used to make the vehicle move and give it speed. Prohibition, which is especially characteristic to the idiomized expression *jumala pärast*, is also compatible with the corresponding meaning 'because of God' in a religious context. *Jumala eest*, on the other hand, has moved farther from the original meanings, which is also attested by the more frequent use in more varied contexts[4].

CONCLUSIONS

In this article we examined the meanings of the Estonian word *jumal* 'god'. First, we presented an overview of the origin of the word *jumal* 'god' and its use in the dialects of Estonian. We then examined the verbs associated with the word *jumal* 'god' based on the data from the corpora of written Estonian. The comparison of frequency data and collocational strength revealed the verbs most frequently used in combination with the word *jumal* 'god'. These verbs can be neutral, having a very general meaning (*tegema* 'do, make', *pidama* 'have to', *võtma* 'take') or specific to God as an agent (e.g., *armastama* 'love', *õnnistama* 'bless'). In combination with some verbs the phrase has acquired holistic interpretations, meaning 'indefinite' or 'random' (e.g., *jumal teab* 'god knows'). *Jumal* 'God' as an interjection is frequently desematicized, exclamatory utterances containing the word *jumal* 'god' express some emotion. These emotional units can be used in different conversational contexts, e.g. for marking direct speech. However, the connection between these phrases and applying to a higher power for assistance is transparent to some extent. The genitive form of the word *jumal* 'god' has grammaticalized into an intensifying adverb in the meaning of 'very' and primarily modifies scalar adjectives (e.g., *jumala ilus* 'very beautiful', lit. 'God's beautiful'). In the final part of the article we examined two synonymous idiomatic adpositional phrases containing the word *jumal* 'god' – *jumala eest* and *jumala pärast* 'for God's sake'. The corpus study showed that *jumala eest* is more frequent in Estonian and the meaning of the phrase may have desemanticized into the meaning 'indeed' (occurring also in combination with a verb in the affirmative of indicative mood). The phrase *jumala pärast*

is less frequent and its usage is restricted mainly to the negative of imperative and conditional moods. It can be observed that while these adpositional phrases have retained a certain semantic connection to the meaning of the same phrases in religious context, both have developed into phraseological units.

In conclusion, it can be said that while speakers are often not referring to a higher power when using the word *jumal* 'god' in modern Estonian, the language use of Estonians is nevertheless based on a Christian background (see also Riistan 2006).

ABBREVIATIONS FOR PARISHES

Har	Hargla
IisR	Coastal Iisaku
Jäm	Jämaja
Kaa	Kaarma
Khk	Kihelkonna
Kod	Kodavere
Krk	Karksi
Kse	Karuse
Lai	Laiuse
Lüg	Lüganuse
Mar	Martna
Mus	Mustjala
Pöi	Pöide
Rei	Reigi
Rõu	Rõuge
Se	Setu

NOTES

[1] http://www2.keeleveeb.ee/dict/corpus/ettenten/about.html.

[2] As such, the expression is somewhat close to the phrase *mine tea* (go.IMPER + know.IMPER), meaning 'do not know') (see Tragel 2003).

[3] For scales see, for example, Sutrop 2004b.

[4] Cf., for example, Geoffrey C. Williams who has shown how the collocations of the word *God* vary in the New Testament, the works of Shakespeare and BNC depending on whether the word *God* refers to an actor who is present (most frequent in NT), an external force (most frequent in the works of Shakespeare), or whether the resonance with the divide meaning has been lost (frequent in BNC) (Williams 2008).

REFERENCES

Arnovick, Leslie K. 2000. *Diachronic Pragmatics: Seven Case Studies in English Illocutionary Development*. Amsterdam & Philadelphia: John Benjamins Publishing Company. DOI: 10.1075/pbns.68.

EES 2012 = *Eesti etümoloogiasõnaraamat.* [Dictionary of Estonian Etymology.] Metsmägi, Iris & Sedrik, Meeli & Soosaar, Sven-Erik (comps., eds.). Tallinn: Eesti Keele Sihtasutus. Available at http://www.eki.ee/dict/ety/index.cgi?Q=l%C3%A4bi&F=M&C06=et, last accessed on May 22, 2016.

EKG II 1993 = Erelt, Mati & Kasik, Reet & Metslang, Helle & Rajandi, Henno & Ross, Kristiina & Saari, Henn & Tael, Kaja & Vare, Silvi (comps.). *Eesti keele grammatika II: Süntaks.* [Estonian Grammar II: Syntax.] Mati Erelt & Tiiu Erelt & Henn Saari & Ülle Viks (eds.). Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.

EKSS 2009 = *Eesti keele seletav sõnaraamat.* [Estonian Explanatory Dictionary.] Margit Langemets & Mai Tiits & Tiia Valdre & Leidi Veskis & Ülle Viks & Piret Voll (eds.) Available at http://www.eki.ee/dict/ekss/, last accessed on May 22, 2016.

EMS II: 6 = Haak, Anu & Juhkam, Evi & Must, Mari & Neetar, Helmi & Nigol, Salme & Niit, Ellen & Oja, Vilja & Ross, Eevi & Viires, Helmi 1997. *Eesti murrete sõnaraamat.* [Dictionary of the Estonian Dialects.] II: 6. Tallinn: Eesti Keele Instituut.

Habicht, Külli & Penjam, Pille & Tragel, Ilona 2010. Kas *tahtma* tahab abiverbiks? [The Verb *Tahtma* 'Want' as an Auxiliary in Estonian.] *ESUKA – JEFUL*, Vol. 2, pp. 115–146. Available at http://jeful.ut.ee/public/files/habichtjt_esuka2.pdf, last accessed on May 25, 2016.

Hughes, Geoffrey 2006. *An Encyclopedia of Swearing*: *The Social History of Oaths, Profanity, Foul Language, and Ethnic Slurs in the English-Speaking World.* Armonk & London: M.E. Sharpe.

Karulis, Konstantīns 1992. *Latviešu etimoloģijas vārdnīca I.* [Dictionary of Latvian Etymology.] Rīga: Avots.

Kehayov, Petar 2009. Taboo Intensifiers as Polarity Items: Evidence from Estonian. *STUF: Language Typology and Universals / Sprachtypologie und Universalienforschung*, Vol. 62, Nos. 1–2, pp. 140–164. http://dx.doi.org/10.1524/stuf.2009.0009.

Koivulehto, Jorma 1999. Varhaiset indoeurooppalaiskontaktit: aika ja paikka lainasanojen valossa. [Early Contacts with Indo-Eroupeans: Time and Place of the Light Loan Words.] In: Paul Fogelberg (ed.) *Pohjan poluilla: Suomalaisten juuret nykytutkimuksen mukaan.* [Northern Paths: Current Research of Finnish Roots.] Helsinki: Societas Scientiarum Fennica, pp. 207–236.

ÕS 2013 = Raadik, Maire & Erelt, Tiiu & Leemets, Tiina & Mäearu, Sirje (eds.) *Eesti õigekeelsussõnaraamat*. [The Spelling Dictionary of Estonian.] Tallinn: Eesti Keele Sihtasutus. Available at http://www.eki.ee/dict/qs/, last accessed on May 22, 2016.

Pajusalu, Renate 2006. *Jumal* ja *kurat* argikeeles. [*God* and *Devil* in Everyday Speech.] In: Enn Kasak & Anne Kull (eds.) *Kuradi tark jumala loll.* Tallinn: Eesti Päevalehe Kirjastus, pp. 17–32.

Riistan, Ain 2006. Teoloogide jumalad. [Gods in Theology.] In: Enn Kasak & Anne Kull (eds.) *Kuradi tark jumala loll.* Tallinn: Eesti Päevalehe Kirjastus, pp. 113–126.

SSA I 1992 = Itkonen, Erkki & Kulonen, Ulla-Maija (eds.) *Suomen sanojen alkuperä: Etymologinen sanakirja 1. A–K.* [Origin of Finnish Words: Etymological Dictionary I. A–K.] Helsinki: Kotimaisten Kielten Tutkimuskeskus & Suomalaisen Kirjallisuuden Seura.

Sutrop, Urmas 2001a. Taarapita – saarlaste suur jumal. [Taarapita – the Great God of the Inhabitants of Saaremaa Island.] *Mäetagused*, Vol. 16, pp. 7–38. http://dx.doi.org/10.7592/MT2001.16.sutrop.

Sutrop, Urmas 2001b. List Task and Cognitive Salience Index. *Field Methods*, Vol. 13, No. 3, pp. 263–276. http://dx.doi.org/10.1177/1525822X0101300303.

Sutrop, Urmas 2003. Eesti keele maailmapildist: meel, hing ja vaim. [On the World View of the Estonian Language: Mind, Soul, and Spirit.] *Mäetagused*, Vol. 24, pp. 99–108. http://dx.doi.org/10.7592/MT2003.24.sutrop.

Sutrop, Urmas 2004a. Oletatav paganlike eestlaste puust jumalakuju Oleviste raamatukogus. [A Wooden God of Pagan Estonians from before 1800?] *Mäetagused*, Vol. 27, pp. 127–134. http://dx.doi.org/10.7592/MT2004.27.sutrop.

Sutrop, Urmas 2004b. Temperatuurisõnad ja skaalad. [Temperature Words and Scales.] In: Mati Erelt (ed.) *Emakeele Seltsi aastaraamat 50.* Tallinn: Emakeele Selts, pp. 111–128. Available at http://www.emakeeleselts.ee/esa/ESA_50_pdf/Sutrop.pdf, last accessed on May 26, 2016.

Tragel, Ilona 2001. On Estonian Core Verbs. In: Ilona Tragel (ed.) *Papers in Estonian Cognitive Linguistics.* Publications of the Department of General Linguistics 2. Tartu: University of Tartu, pp. 145–169. Available at http://dspace.ut.ee/bitstream/handle/10062/41786/papers_in_estonian_ocr.pdf, last accessed on May 26, 2016.

Tragel, Ilona 2003. Eesti keele seriaalkonstruktsioonist. [Serial Constructions in the Estonian Language.] *Keel ja Kirjandus*, No. 12, pp. 919–934. Available at http://www.digar.ee/arhiiv/et/perioodika/35115, last accessed on May 26, 2016.

Tragel, Ilona & Habicht, Külli & Piiroja, Piret 2015. Rise and Fall of the TAKE-Future in Written Estonian. In: Brian Nolan & Gudrun Rawoens & Elke Diedrichsen (eds.) *Causation, Permission, and Transfer: Argument Realisation in GET, TAKE, PUT, GIVE and LET Verbs*. Studies in Language Companion Series 167. Amsterdam: John Benjamins Publishing Company, pp. 353–383. http://dx.doi.org/10.1075/slcs.167.13tra.

Uibo, Udo 2014. *Sõnalood: Etümoloogilisi vesteid.* [Word Stories: Etymological Yarns.] Tallinn: Tänapäev.

Vainik, Ene 2014. *Jumala* jälgi ajamas. [Tracing Back the Word *jumal* ('God').] *Mäetagused*, Vol. 58, pp. 7–34. http://dx.doi.org/10.7592/MT2014.58.vainik.

Veismann, Ann & Tragel, Ilona & Pajusalu, Renate 2002. Eesti keele põhisõnavara operaatoritest: katseid verbide ja kaassõnadega. [Basic Vocabulary Operators in Estonian Language: Testing of Verbs and Adpositions.] In: Haldur Õim & Renate Pajusalu & Tiit Hennoste (eds.) *Tähendusepüüdja: Pühendusteos professor Haldur Õimu 60. sünnipäevaks*. [Catcher of the Meaning: Festschrift for Professor Haldur Õim on the Occasion of the 60th Birthday.] Tartu: Tartu Ülikooli Kirjastus, pp. 312–328.

Williams, Geoffrey C. 2008. The Good Lord and his Works: A Corpus-Driven Study of Collocational Resonance. In: Sylviane Granger & Fanny Meunier (eds.) *Phraseology: An Interdisciplinary Perspective*. Amsterdam & Philadelphia: John Benjamins Publishing Company, pp. 159–173. DOI: 10.1075/z.139.16wil.

http://dx.doi.org/10.7592/FEJF2016.64.khan

THE CAPTIVE KHAN AND THE CLEVER DAUGHTER-IN-LAW

Yuri Berezkin, Evgeny Duvakin

Abstract: The Aarne-Thompson-Uther index contains rich data on the tale repertoire of the main areas of Eurasia and North Africa, but it is still Eurocentric, and does not reflect many widespread tale-types that are not registered in Europe or rarely found across the region. We have revealed several plots of such kind. Their distribution seems to point to the eastern part of the Great Steppe as the area of their origin. Later on waves of nomads might have brought these plots to Europe. In this article, two of them are analysed in detail: *The Encoded Message* (that includes rather different versions) and *The Big Bull*. Both find correspondences in Balto-Finnic traditions, which allows us to discuss them in the context of previously unrecognised or poorly studied parallels between the Caucasus and Northern Europe. The approximate date of these links is the second part of the first millennium A.D.

Keywords: cultural and historical links, Eurasian steppe belt, folklore databases, folktales

It is well known that the Aarne-Thompson-Uther index of *The Types of International Folktales* (Aarne 1910; Aarne & Thompson 1961; Uther 2004) has a whole set of serious drawbacks. The crucial one is not the inconsistence in classification (it is inevitable due to the nature of the material), but combining under the same number of such episodes that are not tied inflexibly to each other and have different areal distribution (Berezkin 2015a: 62–64; 2015b: 154–159). There is no way to ascertain whether a cited narrative of a certain type contains all of the listed episodes and what episodes the text exactly includes if you do not have access to the original publications. It does not mean that the ATU index is quite useless in searching for parallels, but it implies that this system is not as efficient as it could be.

The main reason lies in the Eurocentric character of *The Types of International Folktales*. On the one hand, ATU's power to classify the folklore of Sub-Saharan Africa, Siberia, Southeast Asia, and Oceania is restricted, while Australia and America are completely beyond its scope. On the other hand, even in Nuclear Eurasia where the system is quite working there are many narrative episodes that are still unnoticed and unindexed. As a whole, these

episodes have a wide distribution but specifically in Europe they are not attested or scarce, and therefore have not been numbered in ATU.

The lack of progress in cataloguing international tale-types owes to the fact that the work obtains its importance only if folklore is assumed to be a source of data on the human past; otherwise, this work is of secondary value. So it is not surprising that in the last few decades the challenge of studying the areal distribution of motifs and plots has been issued exclusively by specialists who got into folkloristics from other fields of research.

Elements of culture do not spread itself and do not arise spontaneously (similar ones could emerge in traditions that are historically independent of each other but the similarity is always incomplete and often superficial). The presence of compound texts closely resembling each other at opposite ends of the ecumene is an evidence for links between the people, whether it is migrations or cultural contacts. Our goal is to reveal these links and to examine them in historical perspective.[1]

THE CAPTIVE KHAN AND HIS DAUGHTER-IN-LAW

The collection of folklore records made by Grigory Potanin in South Siberia and Mongolia in the late 1870s contains the following text.

Teleuts. Yeren Chechen wanted his foolish son to marry a clever girl. When it was raining, a girl got undressed, covered firewood with the clothes and her breast with her arms. She explained to Yeren Chechen that her parents are old, and she should kindle a fire just after returning home; her genitals had been seen by her father and mother, but nobody had laid eyes on her breast. Yeren Chechen understood the girl was clever, and married off his son to her. Yara Chechen, to whom Yeren Chechen submitted, had wanted to get his daughter-in-law and so gave him impossible tasks. First of them was to make boots of stone. The daughter-in-law made a heap of sand and said to Yara Chechen that Yeren Chechen has asked her to make a rope of sand. Yara Chechen doubted whether it was a feasible task. The daughter-in-law asked him whether it was possible to make boots of stone. Afterwards Yara Chechen gave to Yeren Chechen a bull and said: "I will visit you after three days. Let the bull calve at the moment; you will entertain me with his milk". After three days, the daughter-in-law started to pluck and heap grass. She said to Yara Chechen that her father-in-law would give birth to a child, so one should spread grass under him. Yara Chechen doubted that a man could do it. The daughter-in-law asked whether it was possible that a bull would calve. Afterwards Yara Chechen ordered Yeren Chechen to visit him neither in a shirt nor naked, neither on a horse nor

afoot, neither going on a road nor going without a road, neither entering his yurt nor staying outside it. The daughter-in-law explained to Yeren Chechen how to cope with the task: he should wrap himself with a net, ride a stick, go betwixt the furrows and stay in the doorway with one leg outside and another one inside the yurt. After Yeren Chechen did it, Yara Chechen threw him into a pit and sent seven messengers to his son. They delivered the letter written by the prisoner: "I am lying on six layers of white silk, covering myself with seven silk blankets, playing checkers and drinking sweet white wine. I have seven rams at my home. You must kill six of them, put out an eye of the seventh, fracture his arm and send him back. Let the daughter-in-law read the letter". After reading the text, the son of Yeren Chechen asserted that his father lived well. The daughter-in-law explained: "The wine means tears, the silk is snow, the checkers are stars, the rams are the messengers". The son killed six of the messengers and put out an eye of the seventh. The daughter-in-law assembled troops and the survived messenger led them to Yara Chechen. He was killed, and Yeren Chechen was released (Potanin 1883, no. 100: 362–363).

The first part of the text is a variant of ATU 875 *The Clever Farmgirl*, but the second one does not have such a correspondence. It is based on the plot that is spread from Eastern Siberia to the Balkans and could be defined in the following manner: *After an extensive search, a powerful man finds a clever wife for his son. Being captured by enemies, he sends with some of them an allegorical letter or oral message for his daughter-in-law. She understands the real meaning of the text, destroys the enemies and releases her father-in-law.*

In Asia the plot is also attested among the northern Yakuts (Illarionov et al. 2008, no. 36: 358–363; Ksenofontov 1992, no. 27: 283–285), western Buryats (Bardakhanova & Gympilova 2008, nos. 24–25, 27–28, 32: 122–129, 130–137, 153–154; Eliasov 1959: 7–19; Khangalov & Zatopliaev 1889, no. 14: 86–92), Khalkha and Ordos Mongols (Dulam & Molomjamts 2010: 37–38; Mikhailov 1962: 78–86; Potanin 1893, no. 19: 178–179), different groups of the Tuvinians (Samdan 1994, no. 9: 307–321; Taube 1994, no. 39: 255–258), the Chelkans (Sadalova 2002, no. 35: 347–355, 427), Kazakhs (Daurenbekov 1979: 193–195), Kirghiz (Brudny & Eshmambetov 1968: 87–97), and Karakalpaks (Volkov & Maiorov 1959: 140–142). In the Caucasus and adjacent territories it is known to the Kalmyks (Yegorov 1978: 19–31), Ossetians (Britaev & Kaloev 1959: 348–356), Kabardians (Alieva & Kardanguschev 1977: 21–26), Abazins (Tugov 1985, no. 80: 255–257), and Abkhazians (Bgazhba 1983: 238–245). Here are some abstracts of the texts.

Yakuts (Olenyoksky District). An old man left his foolish son at home, went down the river and came into a hut. There lived a girl with her father. When she put food in front of the guest, he asked: "How many ladles did you pour in

my plate?" – "If you said how many steps were made by your deer on the way from home to this place, I would give you the answer," said the girl. The old man understood that she was clever and came the next day with his son. They got married. During a hunt, the old man encountered people from a hostile clan. He was caught and tied to a tilted larch. The enemies decided to kill him by the smoke of the fire. The old man asked to pass on a message to his son: "I am exhausted. I am rolling as a cone and fighting young leaves. When you have heard my words, cut down the tops of two birch trees growing right in the north, and then fix your eyes on the west. There will probably be a pine forest with countless trees. Cut down tops of all of them and bring them to me. If you do not know how to cut them, the white stone lying under my bed will help you. If you could not understand these words, the sharp knife lying under your pillow will assist you". The enemy leader sent two warriors to the home of the old man. They passed on the message to his son and daughter-in-law. She explained to him: "*I am exhausted. I am rolling as a cone and fighting young leaves* means that your father is tied by enemies to a tilted tree. *Cut down the tops of two birch trees growing right in the north* means that you must behead the warriors who came to our home. *Fasten your eyes on the west. There will probably be a pine forest with countless trees. Cut down tops of all of them* means that you must destroy all enemy forces. *The white stone lying under my bed* is your father's sword. *The sharp knife lying under your pillow* is me, your clever wife". The son snatched out the sword, beheaded the messengers, killed all the enemies and saved his father (Illarionov et al. 2008, no. 36: 358–363).

Kirghiz. During a journey, Zheerenche-chechen asked his son to cook the food on the horse's mane and to shorten the way. The son set the mane on fire and tried to level the road by a hoe. Zheerenche-chechen explained that he ought to have taken out the food and eaten it, having not dismounted the horse, as well as to have started an interesting conversation. Thereafter, Zheerenche-chechen set out in search of a clever wife for the son. He met girls who carried off handfuls of brushwood. When the rain started, they all scattered and only Karachach covered brushwood with her robe. Zheerenche-chechen married her off to his son. After awhile Zhanybek-khan ordered Zheerenche-chechen to visit him neither afoot nor on a horse, neither going on a road nor going on a field. Karachach advised him to ride a goat and to go between two roads. Zhanybek-khan told Zheerenche-chechen to pull out sinews from a stone and to make boots of sand. Karachach, under the guise of another woman, asked the khan the same thing, and he said that it was impossible to accomplish. Thereafter, Zhanybek-khan captured Zheerenche-chechen and his son. Zheerenche-chechen sent a letter home, saying that crops had to be reaped before the beginning of the winter, the black blind goat had to be thrown into a lake, and the goat with

straight horns had to become the leader. He also wrote that he had a sharp knife and a worthless one, and said not to forget the latter nevertheless. Karachach clarified: the enemies must be killed, Zhanybek-khan must be thrown into a lake, his place must be taken by his younger brother Berdibek, the sharp knife is she, the daughter-in-law, and the worthless knife is her husband. All this was done, and Zheerenche-chechen was released (Brudny & Eshmambetov 1968: 87–97).

Ossetians. An *aldar* (noble) married his son to a daughter of another aldar, and wanted to test her. During a journey, he asked the son to shorten the road. The youth did not understand the meaning of the request, and so the father beat him. At home the daughter-in-law took pity on her husband. The aldar heard these words and expelled her. The same situation was repeated with the new daughter-in-law after the aldar had asked his son to boil a caldron without water. Then the aldar ordered the son to sell a ram and to bring back both its halves. The girl told her old mother to buy the ram, to slaughter and eat it, and to split its head into two halves and give them back to the youth (as it is the share of old men). The aldar was pleased and married his son to the girl. When he beat him once again – this time for misunderstanding the order "to make horses bold" – the new wife explained that it was necessary to let the horses graze. She also clarified the previous requests: in the first case, the youth ought to have started a conversation, while in the second one he ought to have lit the father's pipe. Later on, the aldar was captured by enemies and thrown into a pit. He suggested to hold him for ransom and to send a message to his home. It contained the following directions: bring in 18 x 18 one-horned bulls and as many two-horned ones, take knives away from the wall and sharpen them on an outdoor grinding stone, cut down two of three gold poles, keep the third one safe and bring it here with cattle. The daughter-in-law explained that the one-horned bulls were spearmen, two-horned ones were warriors with shields and swords, knives were the people of the aldar, and the grinding stone was she, the daughter-in-law. She also clarified that two of the envoys had to be killed, and the third one would show the way. The enemies were destroyed and the aldar was released (Britaev & Kaloev 1959: 348–356).

Besides Siberia, Central Asia, and the North Caucasus, the same plot is attested among the Bulgarians. We have found it in Ukrainian translation. In the highly detailed Bulgarian tale-type index (Daskalova-Perkovska et al. 1994) this plot is not reflected.

Bulgarians. A wealthy man gave his son three hundred sheep, but did not give money for their maintenance and said to bring them back alive and healthy after wintering. A girl advised the youth to borrow money and to feed the sheep till the spring, then to shear wool, sell it, and repay the debt as well as the cattle. The wealthy man learned about this and went to make a proposal to the girl to

marry his son. When he asked her about her father, she said to him: "He went to the mill. If he goes around the forest, he will come back soon, and if he goes straight, he will be delayed". Shortly after the girl's father came home. He said that he would give his daughter, but demanded to bring matchmakers. When they came and sat at the table, a dog began to bark outside. The father-in-law went out to see who is coming and, having returned, said: "A *neroda* goes. He leads a *neploda* and carries *nenitca*".[2] The girl clarified: "A monk went. He led a mule and carried salt". When the newly married couple set out, the father-in-law told them to wait for him at straight furrows and dry stumps emitting water. During the journey, the girl explained that they ought to wait near the vineyard. One day, the father-in-law was captured by robbers. He said to them: "Go to my home and tell my daughter-in-law to feed you and give you something to drink. Let her lead you into the cellar where money is kept. You will be able to take as much as you want. I also ask you to communicate to my daughter-in-law the following words: our dog has brought puppies and our sow has farrowed, so take the puppies and throw them under the sow and put the piglets under the dog". The daughter-in-law listened to the message, fed the robbers and gave them to drink, and then said: "Here's the piglets under the dog and the puppies under the pig!" Peasants killed the robbers and released the father-in-law (Ketkov 1979: 151–153).

The records that were retold here and other texts based on the plot contain a whole set of similar episodes: a man asks his son to shorten the way (i.e. to start a conversation) and/or to sell a sheep and to bring back both the animal and the value (a girl teaches the youth how to cope with the task); when the rain starts, girls scatter and one of them covers firewood with her clothes (a man seeking a wife for his son sees that she is clever); all or a few envoys sent by enemies are killed; when naming certain species of domestic animals the captive has in mind certain people.

The texts are not only obviously resembling each other but also attested on territories with similar natural environment. All of these records, with the exception of the Yakut tale, are taken in the Eurasian steppe belt and adjoining mountainous areas, from Lake Baikal to the Balkans (Figure 1.1). In the following, we will call this set of texts Variant 1 (*The Captive Khan and His Daughter-in-Law*). There are also other groups of tales as well as specific single records that share some similar traits with Variant 1 and are dealing with one common theme. It could be named *The Encoded Message* and defined in such a way: *One of the members of a family gets into trouble and sends to the relatives a text or an object (father-in-law to daughter-in-law, daughter-in-law to father-in-law, husband to wife). Only the receiver understands the real meaning of the message, saves the sender and/or destroys the enemies.*

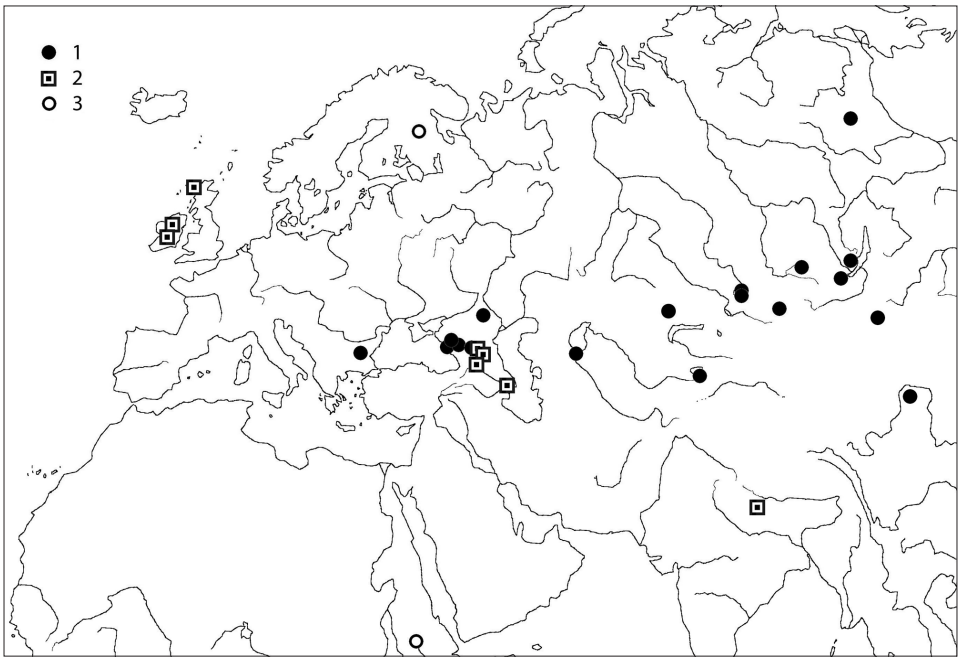

Figure 1. *Distribution of "The Encoded Message". 1. "The Captive Khan and His Daughter-in-Law". 2. "The Crooked Tower and the Builder's Daughter-in-Law". 3. Karelian and Amharan versions related to Variant 1.*

THE CROOKED TOWER AND THE BUILDER'S DAUGHTER-IN-LAW

Variant 2 is close to Variant 1 but attested in other areas (Figure 1.2). We will call it *The Crooked Tower and the Builder's Daughter-in-Law*. A master builder gets to know that a king is going to kill or maim him. He asks the king to send a messenger to his home to bring some ostensibly forgotten objects, usually a special tool. His daughter-in-law or wife understands the real meaning of the message, takes the envoy – in most of versions it is the king's son – hostage and saves the builder.

This variant is very popular among the Irish (Davies 1859: 106–109; Greene 1909: 172–174; Gregory 1905–1906: 75–76; Kelly 1999: 23–27; Kennedy 1866: 68–72; Mac Gréine 1930, no. 4: 262–263; Ó Cianáin 1933: 164–166) and Gaels, including Gaelic-speaking communities of the Outer Hebrides and Cape Breton (MacDougall 1957; MacIntyre 1966; MacLellan 1959; MacNeil 1987, no. 15: 58–60, 243–244).[3] The main figure of all these stories is the legendary carpenter and architect Gobán (or Boban) Saor.

At the same time, the plot is registered in South Asia among the Nepalese (Sakya & Griffith 1980: 127–129) and in the Caucasus among the Ingush (Malsagov 1983, no. 58: 208–210), Chechens (Malsagov 1983: 345; Rossikova 1896, no. 3: 220–222), Georgians (Chikovani 1954, no. 66: 325–327; Kurdovanidze 1988, no. 110: 194–195), and Azerbaijanis (Seidov 1983: 139–144). In the Azerbaijani version, the shah sends to the builder's wife a messenger, a vizier, and, eventually, goes to her personally.

The vast geographical scattering (from the British Isles to Nepal) indicates that some of the texts could remain unknown to us. Nevertheless, even those records that are the most remote from each other have much in common. In the North Caucasus, the border between Variant 1 and Variant 2 separates, on the one hand, speakers of the Northwest Caucasian languages and the Ossetians, and, on the other, speakers of the Northeast Caucasian, namely the Nakh, languages, the Georgians and the Azerbaijanis. Here are some abstracts of these tales.

Gaels (North Uist). Boban Saor sent his son to the market with wethers. He had to get a shilling for them and the wethers back. They were not sold but, on the way back, a girl took them and sheared them, and then gave them back. Boban advised his son to marry this girl, which he did. Then Boban and his son went to Ireland to build a castle and the son's wife advised him to make friends with a girl there. They built the castle but the son's girlfriend told him they were planning to kill them instead of paying them. Boban advised the lord that there was a crack in the wall that could not be fixed without a special tool that he had at home. The lord refused to let them go but allowed his own son to go. Boban's wife understood the message and kept him prisoner till Boban and his son were paid and allowed home safely (MacDougall 1957).

Irish (County Longford). A king wanted to build a tower, so he sent for Gobán Saor. Gobán came with his son and began the work. When the tower was nearly completed, Gobán got a hint that the king was going to make him and his son prisoners, for the fear that anyone else would have another tower like his. Gobán told him that he wanted a special tool to finish the tower, and that his wife wouldn't give it to anyone but himself or his son, for it was a secret. The king refused to let them go but sent his own son. Gobán told the prince to ask his wife for "the crooked and the straight", and to be very careful when bringing it back with him. The prince got to Gobán's house and passed on the message to his wife. The wife said that she would get the tool, and went to a big chest and raised up the lid. She asked the prince to help her. He stuck his head into the chest, and when he did it, Gobán's wife upended him into it and clapped down the lid. Then she sent word to Gobán that she had the prince, and when Gobán went and told the king how he had put him in a fix, he had to let him go (Mac Gréine 1930, no. 4: 262–263).

Ingush. During a journey, an old builder asked his son to shorten the way and to bring a horse. The son did not understand and failed to respond to the requests. "So let's go back home. When we get there, you have to divorce your wife," said the old man. The situation was repeated several times. At long last, the builder's son married a clever girl. She explained that "to shorten the way" means to start a conversation, and "to bring a horse" means to give a staff. Some time later the old man and his son began to build a tower for a king. The builder forefelt that the king would want to kill them instead of paying them, so he had asked his son to start a relationship with the king's daughter. When the tower was almost ready, she alarmed the builders that the king was actually going to kill them. The old man continued the work and intentionally made the top of the tower crooked. Then he told the king that the walls could be straightened by a lever that he had at home. The king sent his son and a servant to fetch it. The builder asked them to pass on a message to his daughter-in-law: "The top of the tower we are building is crooked. Send a lever for straightening towers. Let the crow out of the house, and keep the falcon locked up until our return". The daughter-in-law understood the message and explained its meaning to the relatives. They imprisoned the prince, and sent the servant back. The king had to pay the builders in full and let them go. After they got home, the prince was released (Malsagov 1983, no. 58: 208–210).

Nepalese. A carpenter asked his youngest son to kill a sheep and to bring back goods on it. A girl explained that the father meant to sell the sheep at the market and to buy goods with the money. The carpenter married his son to that clever girl. When the king undertook the construction of a great temple, the father with his three sons went there to find jobs. On the way, he asked them to cut down the mountain. All his sons were shocked and he told them to return home. Only the wife of the youngest son was able to explain that the father meant to tell stories. The next day, the carpenter took his sons to the same spot and repeated the request. The youngest started to tell a story. They arrived in the next country and built the most beautiful pagoda. The king decided to cut off the builders' hands in order that such a temple would never again be built anywhere. The carpenter told the king that they had sent their hands home and that his youngest daughter-in-law was storing those skilled hands there. The king sent his men to fetch the hands but the daughter-in-law told them that she could give them only to the king's son. When the prince arrived, she locked him in a room deep in the basement and sent a messenger to the king: if he cut off any of those men's hands, she will cut off the hands of his son. The king let the builders go back to their home with the rightful reward and the prince returned unharmed (Sakya & Griffith 1980: 127–129).

RARE VERSIONS

Among rare versions of *The Encoded Message*, the most significant are Karelian and Amharic tales (Figure 1.3). Both are closer to Variant 1 than to Variant 2, but, at the same time, have some peculiar traits. In the Karelian tale, not the daughter-in-law saves her father-in-law, but the father-in-law rescues the wife of his son. In the Amharic record, the woman understands the allegorical message and kills the robbers, but they have time to murder her husband, so she finds only his remains.

Karelians (Kalevalsky District). A woman told her son to find a bride in the house where the stove would be lighted earlier than in the rest. The youth came into the house and saw a girl dressed in a chemise and engaged in weaving. She asked him why he had neither mouth nor nose, and told him to tie the horse to winter or summer (they are both in the yard). She also said that her father went for heat, the mother was sick with a last year's merriment, and the brother was going back and forth. The youth did not understand her words, returned home and began to complain that the girl was beautiful but stupid. The father explained that before entering the house he ought to cough; there were a cart and a sledge in the yard; the girl's father went for firewood, her mother gave birth to a child, and the brother was ploughing. After that, the youth married this girl. His father took him into the forest and, when they were going back home, said to him: "Cut off my head and hands, and also throw sticks on the road in order to shorten the way". The son thought the old man had gone out of his mind. At home, he told his wife about it; she explained that he ought to have taken the father's cap and mittens, as well as to have started a conversation. After learning about the girl's cleverness, a king took her away. The old man sent a message to her, saying: "Do you dress well, or do you sometimes wear ragged clothes?" – "On Saturday evenings, the pinafore is ragged," replied the girl. The old man understood that on these evenings the back gate is opened, went there and returned his daughter-in-law (Konkka 1963, no. 58: 374–378).

Amhara. (We omit initial episodes here as they are not directly related to the subject matter). A man together with a rich man were walking along the road. When they were climbing a hill or coming down it, he proposed to the rich man to carry each other in turn. Furthermore, he called small villages with many people "cities", while large but lonely ones "small hamlets". When the rich man got home, he told his relatives about the strange companion. The daughter explained that the companion had proposed to tell stories, and that a village was a town if there was a judge and a mayor. The rich man invited him into his home. The companion demanded that the rich man's daughter become his wife and asked her to sow, grow, and process cotton, and to make clothes from it during a day. The girl told him to grow chickpea and to prepare

oil from it during the same time period. After that they got married. The man left a note and disappeared. There were only two words there – "seven" and "hundred". The girl told her parents that it meant "in the seventh month, I will come with my relatives. There will be one hundred people". After returning, the man carried gold from one place to another and was captured by robbers. They began to consult what was best to be done, to kill him or let him go. A robber that was toothless proposed to leave him alive, but the others decided to slay him. Before his death, the man advised the robbers to visit his home (he had some good wine there) and asked them to say to his wife: "There are seven jugs of wine. Keep the one that has jagged edges, and put the others before them". The wife understood the message and called people for help. The robbers were captured, six of them were killed, the toothless was left alive and told to show the place where the body of the dead man lay (Gankin 1979, no. 116: 159–163).

There are also texts that contain only some episodes of *The Encoded Message*. One such example can be found in early Arabic literature. The captive there sends an allegorical message not in order to save himself but in order to rescue people of his tribe. A similar situation occurs in Yakut historical legends.

Arabs (*Days of the Arabs*, second half of the first millennium A.D.). Warriors of the Banu Bakr b. Wa'il tribe decided to attack people of the Banu Tamim tribe. A man named One-eye who was a member of the Banu Tamim and was at that time a captive of the Banu Bakr had learned about this and asked to send a message to his relatives. A youth was brought to him. One-eye took some sand in his hand and asked: "How much sand is there in my hand?" "There is an innumerable multitude there," replied the youth. One-eye pointed at the sun and asked one more question: "What is it?" – "The sun," was the answer. "You are a quick-witted fellow," said One-eye, "go to my relatives and give them my regards. Tell them to treat their captive from the Banu Bakr well, as the people here are treating me like this. Tell them also to unsaddle my red camel and to ride my brown female camel, remembering that I ate *hais* with them. They should take care of my affairs among Malik's sons. Let them know that blackthorn was covered with leaves, and women made bota bags. They should listen to Khuzail b. al-Akhnas". The youth delivered the message. One-eye's relatives did not understand his words and thought that he had gone mad. Khuzail explained to them: "Taking the sand in his hand, he wanted to say that an innumerable army is moving towards you. Pointing at the sun, he said that it is clearer than the sun. His red camel is Mount Al-Samman, and he tells you to unsaddle him, that is, to get away from this mountain. His brown female camel is the ad-Dahna desert, and he tells you to take cover there. Speaking about Malik's sons, he wants you to warn them about what he had told you, and hold an alliance with them. The blackthorn covered with leaves means that the enemies armed themselves. The women made bota bags for the march. Talking

about *hais*, he has in mind that several clans have assembled to take part in this campaign against you (as *hais* is made from dates, ghee, and curd)". People of the Banu Tamim tribe went to the ad-Dahna desert, and alarmed the Banu Malik b. Hanzalah (Dolinina & Polosin 1983: 112–114).

Yakuts (Vilyuysky District). Mangan Mekchi and his warriors attacked Yungkeebil Khosuun. Before his death, he said: "Let my old father turn toward the east and shoot an arrow from my bow". The warriors delivered the message. Yungkeebil Khosuun's father surmised its meaning, snatched the cradle with his grandson and threw it into a pit wherefrom the soil was taken. The warriors did not notice this and went away. When the son of Yungkeebil Khosuun grew up, he killed Mangan Mekchi as well as his people (Ergis 1960, no. 23: 109–111).

Sometimes the motif of a clever woman who understands the real meaning of her husband's message is incorporated into the tales based on *The Wife Who Would Not Be Beaten* (ATU 888A). Here are corresponding episodes of such an Indian text.

Uttar Pradesh (data on ethnic affiliation are lacking). A merchant's son announced that he would marry only a woman who would allow him to beat her every day. After the wedding his wife said that she would permit him to beat her if he earned money and bought things for their home. The merchant's son set out on a trading expedition, but during the journey encountered some cheaters and lost all his goods. As a result, he had to get fixed up in a job at a dairy. He sat at a press and squeezed oil out of sesame seeds for days on end. Nevertheless, he wrote to his father that he was glad of his destiny and had accumulated great wealth in his hands. The merchant was very pleased, but the wife of his son understood that everything that was written in the letter was a fiction. She disguised herself as a man and helped her husband to return home (Zograf 1976: 218–223).

KING THE CRAFTSMAN AND THE BASKET MAKER

The ATU index includes *The Basket Maker* tale-type (888A*) and defines it in the following manner: *A man learns a trade (basket-making, painting) in order to support a wife. She is taken away by a sea captain. Years later, she recognises her husband when she sees his baskets displayed.* It seems that such a subvariant is typical only for Europe to the west of the River Don while Central Asian, Caucasian, Near Eastern and North African cases are somewhat different. We will call this second type *King the Craftsman*. It includes the following episodes: *A poor girl agrees to marry a prince or a king only if he learns a craft. He does it, marries the girl and then gets into the hands of robbers. He promises them to produce a valuable object that they could sell for good money. His wife or (in*

rare cases) his father recognises his work or reads secret signs on the object. The captive is released, and the criminals are killed.

This story is registered among different Arabic- and Berber-speaking groups of North Africa (Alarcón y Santón 1913, no. 3: 97–103; Nowak 1969, no. 267: 254–255), the Persian (Marzolph 1984, no. *888B: 161), Tajiks (Amonov & Ulug-zade 1957: 352–354), Uzbeks (Afzalov et al. 1972: 445–447), Anatolian Turks (Eberhard & Boratav 1953, no. 231: 278), Armenians (Buniatov 1898, no. 7: 107–109; Gullakian 1990: 40; Harutyunian 1986: 138–140), Georgians (Chikovani 1954, no. 72: 345–346; Kurdovanidze 1988, no. 118: 227–232), Tats (Kukullu 1974, no. 33: 250–256), Lezgins (Alieva 2013, no. 16: 82–83), Avars (Alieva 2013, no. 55: 250–251), Dargins (Alieva 2013, no. 20: 93–94), Kabardians (Tambiev 1900, no. 1: 1–8), and Volga Tatars (Zamaletdinov 2009, no. 62: 251–255). It is also found in the Wortley-Montague manuscript of the *Arabian Nights* (Marzolph & Leeuwen 2004: 416), which is dated 1764–1765 and comes from Northern Egypt (Moussa-Mahmoud 1976: 7–17). The easternmost versions are recorded among the Kazakhs (Sidelnikov 1964: 116–119) and Dungans (Riftin et al. 1977, no. 49: 233–235) (Figure 2).

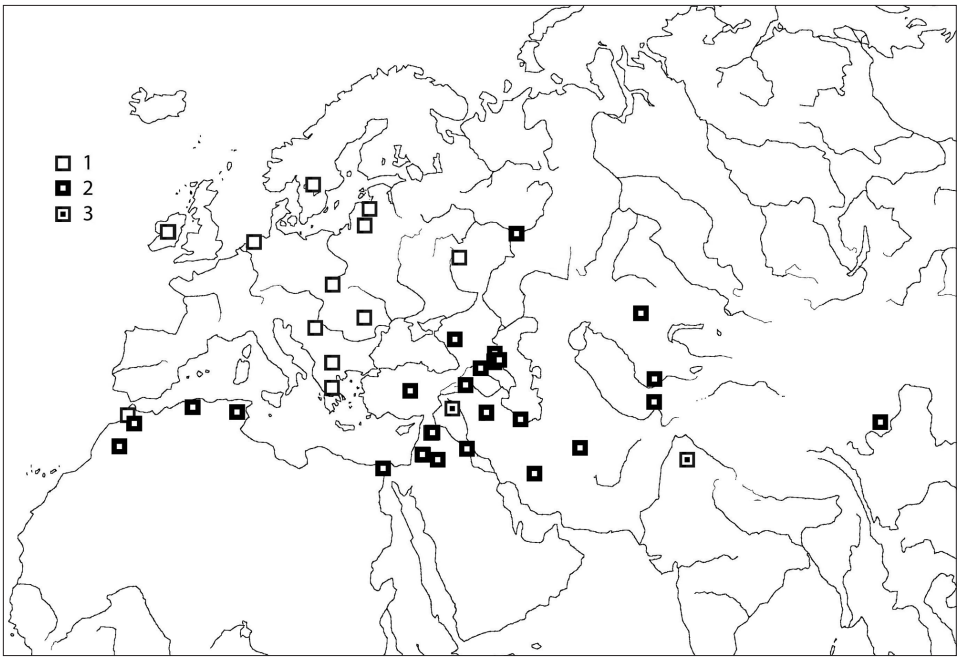

***Figure 2**. Distribution of "King the Craftsman" and "The Basket Maker". 1. "The Basket Maker" (ATU 888A*). 2. "King the Craftsman" (El-Shamy 2004, no. 949*). 3. Either subvariant 1 or 2 (text or detailed enough abstract are unavailable).*

Most of the European tales cited under ATU 888A* still remain unavailable to us. Fortunately, in the Arabian index (El-Shamy 2004), the *King the Craftsman* is selected as a separate tale-type 949* (*Young Gentleman Learns Basketwork. Skill at a craft / trade saves him*). Hasan El-Shamy does not provide the reader with detailed abstracts, but the sets of motifs mentioned by him under this type and type 888A* are different, which allows to classify stories in question with reasonable certainty. It seems that Hans-Jörg Uther was unfamiliar with the corresponding Arabian and other African as well as Asian versions of the *King the Craftsman* and because of this did not pay attention to the difference between two types noticed by El-Shamy. The evidence suggests that *The Basket Maker* is typical mostly for Europe with only one Berberian case in Morocco while the *King the Craftsman* has not been found in Europe to the west of the Don and is typical for Asia and Northern Africa. The first plot is rather remotely related to other versions of *The Encoded Message* while the second demonstrates clear parallels with *The Captive Khan* and *The Crooked Tower*.

The most common variant of the *King the Craftsman* is associated with carpet making. These are Dungan, Uzbek, Tajik, Persian, Volga Tatar, Tat, Lezgian, and Armenian versions, as well as one of the Georgian records.

Dungans (text collected in southeastern Kazakhstan in the dialect spoken in Gansu). A young emperor wanted to marry a poor girl. She refused him as he did not learn any craft. The emperor had mastered the art of carpet making and weaved a wonderful carpet for this girl. They got married. After a while, the emperor disguised himself as a commoner and began to travel from one city to another. During the journey, he came to a tavern. The tavern owner and his people put him in a basement in order to slaughter him for meat. The emperor convinced them that it would be more profitable to let him make an expensive carpet. When the carpet was ready, he told the tavern's owner to sell it to the emperor. The empress understood that the carpet was a work of her husband, found, and released him. After that, all robbers were killed (Riftin et al. 1977, no. 49: 233–235).

Tats. A padishah fell in love with a herder's daughter. She said she would marry him if he learned a craft. The padishah had mastered the art of carpet making and married the girl. One day he walked in the guise of a dervish, entered a tavern, and fell in the basement. The tavern owners were robbers who were selling human flesh. The padishah told them that they would get more money if he wove a carpet. When the carpet was ready, he told them to take it to a certain shop. The padishah's wife saw the carpet and went to the tavern with warriors. The padishah was released, and the robbers were beheaded (Kukullu 1974, no. 33: 250–256).

Volga Tatars. A padishah's son wanted to marry a poor girl. She told him to learn a craft beforehand. The young padishah had mastered the art of carpet making and married the girl. After a while, he set out on a journey through his estates. He visited many places, but one day came to an unknown city and was captured by robbers. He told them to get him to their leader. This was the governor of the city. The young padishah told him that he could weave expensive carpets for sale in the padishah's palace. The governor liked this idea. When the first carpet was delivered in the palace, the wife of the vanished padishah noticed that between its patterns the name of her husband was woven. She bought the carpet and asked to bring her some more. Each following carpet had a new message concerning the padishah's misadventure. In this way, the wife learned where he was and who had captured him. When she received the tenth carpet, she ordered to equip forces. The padishah was released, and other prisoners were permitted to punish the robbers (Zamaletdinov 2009, no. 62: 251–255).

The craft that the hero learns in order to marry may be different than carpet making, but such variants are less frequent. There is, for example, mat weaving (Turkish versions and the tale from the Wortley-Montague manuscript), silk spinning (Arabic versions from North Africa), and hat making (Kazakh version).

Kazakhs. A khan's son, Idris, fell in love with a peasant's daughter called Duriya. When he asked her to marry him, she refused him as he had not learned any craft. Idris returned to his palace and learned hat-making. Duriya agreed to marry him. After the wedding, she asked Idris to make a woollen cap and taught him how to embroider patterns which meant letters and words. Some time later, Idris disguised himself as a beggar and set out on a journey. During the journey, he came in a city, entered a house, and was captured by cannibals who were living there. He said that he could make a beautiful cap and that the khan's wife (by this time Idris was already the khan) would buy this cap for good value. When the cap was ready, two of the cannibals took it to the palace. Upon looking at this cap, Duriya understood what had happened to her husband. Idris was released, and the surviving cannibals were hung (Sidelnikov 1964: 116–119).

THE BIG BULL AND THE DATING OF THE DISCUSSED TALES

The very fact that there are links between traditions in the eastern and western parts of the Eurasian steppe belt is not surprising. The only question is how they are dated.

It is clear that the plots discussed here did not emerge in the Palaeolithic. There are no parallels for them in the New World. *The Captive Khan* is associated with cattle-breeding steppe tradition, while *The Crooked Tower* and the *King the Craftsman* are associated with settled culture. The third plot is, without fail, set in a city, and the second one implies the existence of power centres and permanent fortifications.

The Captive Khan is widespread in the Eurasian steppe belt and adjoining mountainous areas (Altai, central and western parts of the North Caucasus), *The Crooked Tower* is attested in Nepal, South Caucasus, eastern part of the North Caucasus, and on the British Isles, and the area of distribution of *The Basket Maker* and the *King the Craftsman* embraces the main territory of Nuclear Eurasia, within which many dozens of other fairy-tale and novelistic motifs are circulated. Although all these stories can be traced back to a single prototype, they were obviously distributed across different social environments and hardly at the same time.

The eastern origin of *The Captive Khan* is likely simply because, since the middle of the first millennium B.C., all migrations across the Great Steppe were from east to west. Where the other stories branched off from this plot or from their common prototype is hard to say. The motif of the robbers' den, in which people were slaughtered for meat or being squeezed of their fat, is found both in the east (inter alia, in the 14th-century Chinese novel *Water Margin*, and in the Japanese *Anthology of Tales from the Past*, dating back to ca. 1050 A.D.: Ikeda 1971, no. 956A: 214; Riftin 1977: 472; Shi Nai'an & Luo Guanzhong 1981: 924), and in the west of Eurasia (ATU 956). Although the similarity of *The Basket Maker* with *The Captive Khan* is minimal, the Karelian tale has characteristic features of both of them. This may mean that during the formation of the northern version connections between variants of *The Encoded Message* were obvious enough.

In the Mongol-speaking world, *The Captive Khan* is attested among the Ordos, Khalkha, western Buryats, and Kalmyks. Taking into account the origin of the latter, we can assume that the plot is also known to the Oirats. The Kalmyk version shares exclusive parallels with the Khalkha, Buryat, and Ossetian tales. Only in these traditions, the son is trying to cut out a wooden boiler to cook food, after his father has asked (in allegorical form) to light up his pipe or roast meat on a spit.

The existence of the Yakut versions clearly related to the Mongolian tales indicates that the story was known in Inner Asia and South Siberia from at least the 12th–13th centuries, i.e. before the migration of the Yakuts to the lowlands of the middle course of the River Lena. At the end of the 16th – early 17th century, the Kalmyks brought it from Dzungaria to the west, but they were

hardly alone. Among the Kalmyks, *The Captive Khan* is relatively rare, therefore it is unclear why this very plot should have been borrowed by the population of Kazakhstan and the North Caucasus and has become popular there.

It is also not very likely that the initial spread of *The Captive Khan* took place during the period of the Golden Horde. In the Russian-language publications of Bashkir and Volga Tatar folklore, no parallels for this plot have been found. In Dagestan, Variant 1 and Variant 2 are lacking and there are only tales based on the *King the Craftsman*, although this is the area that was connected with the Golden Horde more closely than the rest of the North Caucasus.

In the texts, collected among the Yakuts, Buryats, Khalkha, Tuvinians, and Kirghiz, the captive allegorically calls his daughter-in-law 'knife' or 'scissors'. In the Caucasus a similar comparison ("stone for sharpening knives") is found only in the Ossetian version. This is an argument in favour of the claim that the initial spread of the plot took place in earlier times; namely, the second part of the first millennium A.D., i.e. during the period of intensive interaction between the Iranian-speaking steppe tribes and the Turkic-speaking nomads, moved from the east towards the west, or even as early as the first half of the first millennium B.C., when Scythian groups migrated from Inner Asia to the North Pontic area.

Evidence in favour of this hypothesis is the presence of the geographically isolated Karelian version. It differs significantly from the southern records, but, at the same time, is obviously related to Variant 1, which is typical for the steppe areas. Concerning *The Encoded Message*, the Karelian parallel is unique, but there are many other plots and motifs that link up the Caucasus and the steppe zone, on the one hand, and Northern Europe, on the other. For reasons of space, it is not possible to discuss these data in detail here, so we will give only one example, *The Big Bull* motif (Figure 3).

The ATU tale-type 1960A with a similar name is a conglomerate of any references to domestic and wild animals of enormous size. The unique Olonets bylina tells of a giant bull, which is killed and cut up (Hilferding 1983: 286–291). This motif is also found in Balto-Finnic runic poetry, particularly in the Estonian, Ingrian, Finnish, and Karelian epic (Abercromby 1893, no. 48g: 45–46; Laugaste 1998: 9; Kiuru 1990, no. 8: 47–48; Siig 2013: 22–35; Tedre 1969, no. 4: 17; Yevseev 1950, no. 65: 146). In contrast to the bylina, the Balto-Finnic variants typically contain the motif of determination of the size of the bull by reference to the time during which an animal or bird overcomes the distance from one part of the body to another. A similar motif (the time that it takes for an unmounted man or a horseman to walk or ride from the tail to the middle of the body and from the middle to the head) is attested in the Caucasus and the steppe zone, particularly among the Gagauz, Moldovans, Adyghe, Kabardians,

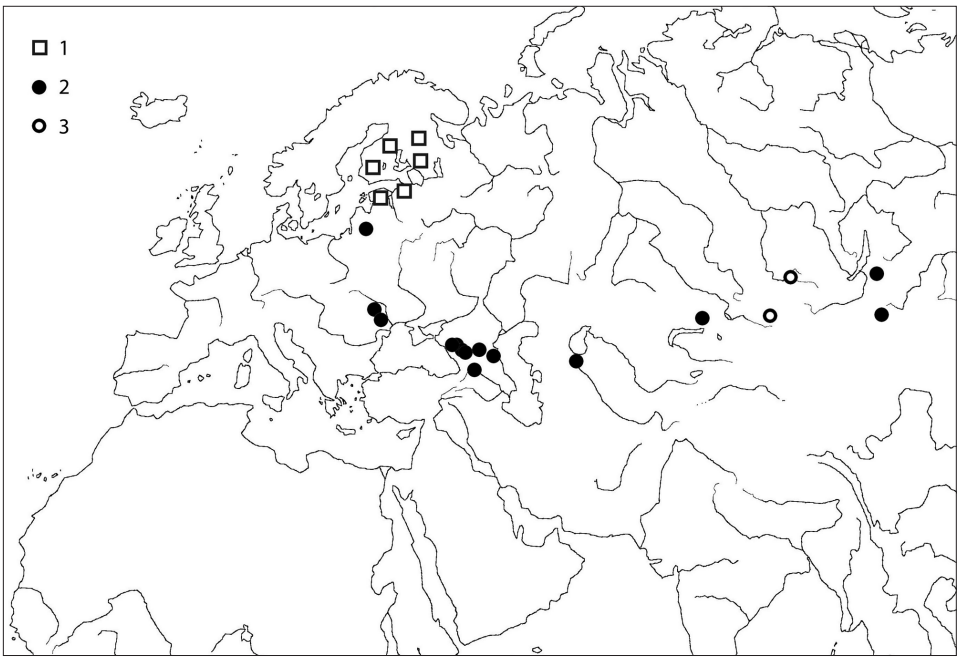

Figure 3. *Distribution of "The Big Bull". 1. Balto-Finnic versions. 2. Typical southern versions. 3. Tuvinian versions.*

Karachays and/or Balkars, Nogais, Georgians, Kazakhs, Karakalpaks, Khorin Buryats and Khalkha Mongols, as well as among the Lithuanians (Alieva 1978, nos. 25–26: 223–227; Alieva & Kardangushev 1977: 123–125; Alieva & Kholaev 1983: 90–97; Benningsen 1912: 18–19; Botezatu 1981: 368–371; Chikovani 1954: 409–410; Kerashev 1957: 271–272; Liobytė 1965: 277–278; Lopatinski 1891: 93–96; Moshkov 1904, no. 136: 204–205; Nogai 1979, no. 18: 86–102; Potanin 1893, no. 4: 378). In the Lithuanian as well as in all steppe and Caucasian versions, this episode is followed by a standard series of other ones: an eagle takes away the bull, sits down on a goat's horns and drops the bull's scapula into an old man's eye, somebody takes this bone out of the eye, it is covered with soil and settled by people, a fox begins to pull the scapula, the people kill the fox, but cannot move her corpse, a woman or a child does it easily. The same series of episodes, but without the motif of determination of the size of the bull by reference to the time spent on the road from one place to another, is found in Abazin, Karachay, Svan, Ossetian and Ingush texts (Aleinikov 1883, no. 4: 155–157; Britaev & Kaloev 1959: 96; Dalgat 1972: 298–299; Kibiev & Malsagov 1981: 26–28; Palmaitis 1986: 134–137; Shanaev 1870: 14–15; Tugov 1985, no. 121: 336–337).

In variants from Tuva, a goat appears instead of a giant bull (Samdan 1994, no. 28: 399). The record made among the Altai Tuvinians is confusing and incomplete. A bull is mentioned along with a traveller, an old man, a bird, and a goat, but the initial episode, which describes how the bull was taken away by the bird, is omitted (Taube 1994, no. 57: 276–277).

In terms of genre all variants from Finland to Mongolia are similar. They can be easily integrated into different contexts, and, taken in isolation, are closely related to cumulative and tall tales. An important common motif is the paradoxical characteristic of beings and objects as simultaneously giant and tiny, mighty and weak. This also applies to most other 'Balto-Caucasian' plots that we do not consider here.

The area of distribution of *The Big Bull* in its southern variant largely coincides with that of *The Captive Khan and His Clever Daughter-in-Law*. One should note that both plots are completely lacking in the Volga region. Based only on the data presented here, it is not possible to determine the exact time of their penetration into Northern Europe, but, as has been said, there are many other parallels as well. Most of them have the distribution, which is broken up into two unconnected areas – Caucasus and the Great Steppe, on the one hand, and the Circum-Baltic region, on the other. It is logical to assume that the transfer of folklore over long distances was most easily carried out during the period of a drastic alteration of the ethnocultural map of both Europe and the Asian steppe, i.e. in the period from Late Antiquity to the Mongol invasion. A more narrow range that may be proposed for the time being is the second part of the first millennium A.D.

ACKNOWLEDGEMENTS

The work is supported by the Russian Science Foundation, project 14–18–03384 ('Stories Retold over Millennia: Reconstruction of the Dynamics of the Global Distribution of Replicated Elements of Oral Narratives').

NOTES

[1] The study is based on the electronic catalogue of world folklore and mythology. On April 7, 2016, the database contained more than 50,000 abstracts of texts providing information on the spread of 2120 motifs according to 925 traditions. Its online version is updated once a year and available (in Russian) at http://www.ruthenia.ru/folklore/berezkin.

[2] *Neroda*, *neploda*, and *nenitca* are dialect names for 'monk', 'mule', and 'salt' correspondingly.

[3] The text collected by Mrs Gomme from an old woman at Deptford and published in Jacobs' *More English Fairy Tales* (1894: 54–55) has more likely a Goidelic origin too.

REFERENCES

Aarne, Antti 1910. *Verzeichnis der Märchentypen*. Helsinki: Academia Scientiarum Fennica. (Folklore Fellows' Communication, Vol. 3).

Aarne, Antti & Thompson, Stith 1961. *The Types of the Folktale: A Classification and Bibliography*. Folklore Fellows' Communication, Vol. 184. Helsinki: Academia Scientiarum Fennica.

Abercromby, John 1893. Magic Songs of the Finns. *Folk-Lore*, Vol. 4, No. 1, pp. 27–49.

Afzalov, Mansur & Rasulev, Kh. & Khusainova, Z. 1972. *Uzbekskie narodnye skazki*. [Uzbek Folktales.] Vol. 1. Tashkent: Izdatel'stvo literatury i iskusstva.

Alarcón y Santón, Maximiliano 1913. *Textos árabes en dialecto vulgar de Larache*. Madrid: Imprenta Ibérica.

Aleinikov, Moisei 1883. Karachaevskie skazaniia. [Karachay Legends and Tales.] *Sbornik materialov dlia opisaniia mestnostei i plemen Kavkaza*, Vol. 3, Part 2, pp. 138–168.

Alieva, Alla 1978. *Skazki adygskikh narodov*. [Folktales of Adyghe Peoples.] Moscow: Nauka.

Alieva, Fatima 2013. *Svod pamiatnikov fol'klora narodov Dagestana*. [Corpus of Folklore Records of the Peoples of Dagestan.] Vol. 3. *Bytovye skazki* [Realistic Tales.] Moscow: Nauka.

Alieva, A. & Kholaev, A. 1983. *Balkarskie i karachaevsckie skazki*. [Balkar and Karachay Folktales.] Moscow: Detskaia literatura.

Alieva, A. & Kardanguschev, Z. 1977. *Kabardinskie narodnye skazki*. [Kabardian Folktales.] Moscow: Detskaia literatura.

Amonov, Rajab & Ulug-zade, Klavdia 1957. *Tadjikskie narodnye skazki*. [Tajik Folktales.] Stalinabad: Tajikgosizdat.

Bardakhanova, Svetlana & Gympilova, Sesegma 2008. *Buriatskie narodnye skazki. Volshebnye. Bytovye*. [Buryat Fairy and Realistic Tales.] Ulan-Ude: Izdatel'stvo Buriatskogo nauchnogo tsentra.

Benningsen, Adam 1912. *Legendy i skazki Tsentralnoi Azii, sobrannye grafom A.P. Benningsenom*. [Legends and Tales of Central Asia Collected by Count A.P. Benningsen.] Saint Petersburg: n.p.

Berezkin, Yuri 2015a. Folklore and Mythology Catalogue: Its Lay-Out and Potential for Research. *The Retrospective Methods Network Newsletter*, Vol. 10, pp. 56–70.

Berezkin, Yuri 2015b. Rasprostranenie fol'klornykh motivov kak obmen informatsiyei, ili Gde zapad granichit s vostokom. [Areal Spread of Folklore Motifs as Information Exchange, or about the Borderline between West and East.] *Forum for Anthropology & Culture*, No. 26, pp. 153–170.

Bgazhba, Khukhut 1983. *Abkhazskie skazki*. [Abkhaz Folktales.] Sukhumi: Alashara.

Botezatu, Grigore 1981. *Moldavskie narodnye skazki*. [Moldavian Folktales.] Kishinev: Literatura artistike.

Britaev, Sozyryko & Kaloev, G. 1959. *Osetinskie narodnye skazki*. [Ossetian Folktales.] Moscow: Gosudarstvennoe izdatel'stvo khudozhestvennoi literatury.

Brudny, Dmitri & Eshmambetov, Kasymbek 1968. *Kirgizskie skazki*. [Kirghiz Folktales.] Moscow: Khudozhestvennaia literatura.

Buniatov, Grigory 1898. Armianskie predaniia i skazki. [Armenian Legends and Tales.] *Sbornik materialov dlia opisaniia mestnostei i plemen Kavkaza*, Vol. 24, Part 2, pp. 96–151.

Chikovani, Mikhail 1954. *Gruzinskie narodnye skazki (sto skazok)*. [Georgian Folktales (One Hundred Tales).] Tbilisi: Zaria Vostoka.

Dalgat, Uzdiat 1972. *Geroicheskii epos chechentsev i ingushei: Issledovanie i teksty*. [Chechen and Ingush Heroic Epic: Study and Texts.] Moscow: Nauka.

Daskalova-Perkovska, Liliana & Dobreva, Dorotea & Kotseva, Yordanka & Mitseva, Yevgenia 1994. *Blgarski folklorni prikazki. Katalog*. [Bulgarian Folktales. The Catalogue.] Sofia: Sv. Kliment Okhridski.

Daurenbekov, Tanash 1979. *Kazakhskie narodnye skazki*. [Kazakh Folktales.] Alma-Ata: Zhalyn.

Davies, Francis Robert 1859. Legends of the County Clare. In: *Choice Notes from "Notes and Queries": Folk Lore*. London: Bell and Daldy, pp. 92–110.

Dolinina, Anna, & Polosin, Vladimir 1983. *Araviiskaia starina. Iz drevnei arabskoi poezii i prozy*. [Arabian Antiquity. Selections from Ancient Arabian Poetry and Prose.] Moscow: Nauka.

Dulam, Sendenjav & Lkhagvasüren, Molomjamts 2010. Mother Tongue, Oral Traditions. In: Urtnasan Norov (ed.). *Intangible Cultural Heritage of the Mongols*. Jeonju: International Information and Networking Centre for Intangible Cultural Heritage in the Asia-Pacific Region under the auspices of UNESCO, pp. 21–43.

Eberhard, Wolfram & Boratav, Pertev Naili 1953. *Typen türkischer Volksmärchen*. Akademie der Wissenschaften und der Literatur. Veröffentlichungen der orientalischen Kommission. Bd. 5. Wiesbaden: Franz Steiner Verlag.

Eliasov, Lazar 1959. *Altan-Khaisha – zolotye nozhnitsy. Buriatskie narodnye skazki*. [Altan-Khaisha – The Gold Scissors. Buryat Folktales.] Moscow: Gosudarstvennoe izdatel'stvo detskoi literatury.

El-Shamy, Hasan M. 2004. *Types of the Folktale in the Arab World: A Demographically Oriented Tale-Type Index*. Bloomington: Indiana University Press.

Ergis, Georgy 1960. *Istoricheskie predaniia i rasskazy yakutov*. [Historical Legends and Stories of the Yakuts.] Vol. 1. Moscow and Leningrad: Izdatel'stvo Akademii nauk SSSR.

Gankin, Emmanuil 1979. *Amkharskie narodnye skazki*. [Amharic Folktales.] Moscow: Nauka.

Greene [Miss Greene] 1909. County Kildare Folk-Tales. *Journal of the County Kildare Archaeological Society*, Vol. 6, No. 2, pp. 172–174.

Gregory [Lady Gregory] 1905–1906. The Goban Saor. *Journal of the Galway Archaeological and Historical Society*, Vol. 4, No. 2, pp. 73–77.

Gullakian, Syuzanna 1990. *Ukazatel' syuzhetov armianskikh volshebnykh i novellisticheskikh skazok*. [A Type Index of Armenian Fairy and Realistic Tales.] Yerevan: Respublikanskii uchebno-nauchnyi kabinet.

Harutyunian, Sargis 1986. *Armianskie narodnye skazki*. [Armenian Folktales.] Yerevan: Sovetakan grokh.

Hilferding, Alexander 1983. *Onezhskie byliny*. [Onega Bylinas.] Arkhangelsk: Severo-zapadnoe knizhnoe izdatel'stvo.

Ikeda, Hiroko 1971. *A Type and Motif Index of Japanese Folk-Literature*. Folklore Fellows' Communication, Vol. 209. Helsinki: Suomalainen Tiedeakatemia.

Illarionov, Vasili & Diakonova, Yulia & Mukhopleva, Svetlana & Nesterova, Lytsia 2008. *Yakutskie narodnye skazki*. [Yakut Folktales.] Pamiatniki fol'klora narodov Sibiri i Dal'nego Vostoka, Vol. 27. Novosibirsk: Nauka.

Jacobs, Joseph 1894. *More English Fairy Tales*. London: David Nutt.

Kelly, Éamon 1999. *Ireland's Master Storyteller: The Collected Stories*. Cork: Mercier Press.

Kennedy, Patrick 1866. *Legendary Fictions of the Irish Celts*. London: Macmillan and Co.

Kerashev, Tembot 1957. *Adygeiskie skazki*. [Adyghe Folktales.] Maykop: Adygeiskoe knizhnoe izdatel'stvo.

Ketkov, Oleksandr 1979. *Bolgars'ki narodni kazki*. [Bulgarian Folktales.] Kiev: Veselka.

Khangalov, Matvei & Zatopliaev, Nikolai 1889. *Buriatskie skazki i poveria*. [Buryat Tales and Beliefs.] Zapiski Vostochno-Sibirskogo otdela Imperatorskogo russkogo geograficheskogo obshchestva po otdeleniyu etnografii, Vol. 1, Part 1. Irkutsk: Vostochnoe obozrenie.

Kibiev, Musbek & Malsagov, Akhmet 1981. *Chechenskie i ingushskie narodnye skazki*. [Chechen and Ingush Folktales.] Moscow: Detskaia literatura.

Kiuru, Eino 1990. *Ingermanlandskaia epicheskaia poeziia*. [Ingrian Epic Poetry.] Petrozavodsk: Karelia.

Konkka, Unelma 1963. *Karel'skie narodnye skazki*. [Karelian Folktales.] Moscow & Leningrad: Izdatel'stvo Akademii nauk SSSR.

Ksenofontov, Gavriil 1992. *Uraangkhai-sakhalar: Ocherki po drevnei istorii yakutov*. [The Uraangkhai-Sakhalar: Essays on the Ancient History of the Yakuts.] Vol. 1, Book 2. Yakutsk: Natcional'noe knizhnoe izdatel'stvo Respubliki Sakha.

Kukullu, Amaldan 1974. *Zolotoi sunduk (skazki tatov Dagestana)*. [The Gold Chest (Folktales of the Dagestanian Tats).] Moscow: Nauka.

Kurdovanidze, Teimuraz 1988. *Gruzinskie narodnye skazki*. [Georgian Folktales.] Book 2. Moscow: Nauka.

Laugaste, Eduard 1998. *Suur Tamm. Allikmaterjale iseseisvaks tööks eesti keele ja kirjanduse eriala üliõpilastele*. [The Big Oak.] Tartu: Tartu Riikliku Ülikooli eesti kirjanduse ja rahvaluule kateeder.

Liobytė, Aldona 1965. *Koroleva Lebed'. Litovskie narodnye skazki*. [The Swan Queen. Lithuanian Folktales.] Vilnius: Vago.

Lopatinski, Lev 1891. Kabardinskie predaniia, skazaniia i skazki, zapisannye po-russki. [Kabardian Legends and Tales Recorded in Russian.] *Sbornik materialov dlia opisaniia mestnostei i plemen Kavkaza*, Vol. 12, Part 1, pp. 1–144.

MacDougall, Donald 1957. Boban Saor and His Son Saved as a Result of the Son's Wife's Advice. *School of Scottish Studies Archive SA1957.080, Tobar an Dualchais / Kist o Riches*. Available at http://www.tobarandualchais.co.uk/fullrecord/42721/1, last accessed on January 22, 2016.

Mac Gréine, Pádraig 1930. A Longford Miscellany. *Béaloideas: The Journal of the Folklore of Ireland Society*, Vol. 2, No. 3, pp. 261–273.

MacIntyre, Michael 1966. Boban Saor and How He Uncovered a Plan to Have Him Hanged. *School of Scottish Studies Archive SA1966.084, Tobar an Dualchais / Kist o Riches*. Available at http://www.tobarandualchais.co.uk/fullrecord/60776/1, last accessed on January 22, 2016.

MacLellan, Angus 1959. How Boban Saor and His Son Got Back Safely from France with the Help of His Son's Wife. *School of Scottish Studies Archive SA1959.057, Tobar an Dualchais / Kist o Riches*. Available at http://www.tobarandualchais.co.uk/fullrecord/41005/1, last accessed on January 22, 2016.

MacNeil, Joe Neil 1987. *Tales until Dawn: The World of a Cape Breton Gaelic Story-Teller*. Translated and edited by John Shaw. Kingston and Montreal: McGill-Queen's University Press.

Malsagov, Akhmet 1983. *Skazki i legendy ingushei i chechentsev*. [Ingush and Chechen Folktales and Legends.] Moscow: Nauka.

Marzolph, Ulrich 1984. *Typologie des persischen Volksmärchens*. Beiruter Texte und Studien. Bd. 31. Beirut: Orient-Institut der Deutschen Morgenländischen Gesellschaft.

Marzolph, Ulrich & Leeuwen, Richard van 2004. *The Arabian Nights Encyclopedia*. Santa Barbara, Denver & Oxford: ABC-CLIO.

Mikhailov, Georgy 1962. *Mongol'skie skazki*. [Mongolian Folktales.] Moscow: Khudozhestvennaia literatura.

Moshkov, Valentin 1904. *Narechiia bessarabskikh gagauzov*. [Dialects of the Bessarabian Gagauz.] Saint Petersburg: Imperatorskaia akademiia nauk.

Moussa-Mahmoud, Fatma 1976. A Manuscript Translation of the *Arabian Nights* in the Beckford Papers. *Journal of Arabic Literature*, Vol. 7, pp. 7–23.

Nogai, Azhdaut 1979. *Nogaiskie narodnye skazki*. [Nogai Folktales.] Moscow: Nauka.

Nowak, Ursula 1969. *Beiträge zur Typologie des arabischen Volksmärchens*. Inaugural-Dissertation zur Erlangung der Doktorwürde der Philosophischen Fakultät der Albert-Ludwig-Universität zu Freiburg im Breisgau. Freiburg im Breisgau.

Ó Cianáin, S.F. 1933. Folk Tales from Co. Sligo. *Béaloideas: The Journal of the Folklore of Ireland Society*, Vol. 4, No. 2, pp. 164–167.

Palmaitis, Mykolas 1986. *Upper Svan: Grammar and Texts*. Vilnius: Mokslas.

Potanin, Grigory 1883. *Ocherki severo-zapadnoi Mongolii*. [Essays on Northwestern Mongolia.] Vol. 4. Materialy etnograficheskie. [Ethnographic Materials.] Saint Petersburg: V. Kirshbaum.

Potanin, Grigory 1893. *Tangutsko-tibetskaia okraina Kitaia i Tsentral'noi Mongolii*. [The Tangut-Tibet Border of China and Central Mongolia.] Vol. 2. Saint Petersburg: A.S. Suvorin.

Riftin, Boris & Khasanov, M. & Yusupov, I. 1977. *Dunganskie narodnye skazki i predaniia*. [Dungan Folktales and Legends.] Moscow: Nauka.

Rossikova, Anna 1896. Puteshestvie po tcentral'noi chasti gornoi Chechni. [A Journey through the Central Part of Mountainous Chechnya.] *Zapiski Kavkazskogo otdela Imperatorskogo russkogo geograficheskogo obshchestva*, Vol. 18, pp. 139–228.

Sadalova, Tamara 2002. *Altaiskie narodnye skazki*. [Altai Folktales.] Pamiatniki fol'klora narodov Sibiri i Dal'nego Vostoka, Vol. 21. Novosibirsk: Nauka.

Sakya, Karna & Griffith, Linda 1980. *Tales from Kathmandu. Folktales from the Himalayan Kingdom of Nepal*. Brisbane: House of Kathmandu.

Samdan, Zoia 1994. *Tuvinskie narodnye skazki*. [Tuvinian Folktales.] Pamiatniki fol'klora narodov Sibiri i Dal'nego Vostoka, Vol. 8. Novosibirsk: Nauka.

Seidov, Nureddin 1983. *Azerbaidzhanskie skazki*. [Azerbaijan Folktales.] Baku: Maarif.

Shanaev, Dzhantemir 1870. Osetinskie narodnye skazaniia. [Ossetian Folktales and Legends.] *Sbornik svedenii o kavkazskikh gortsakh*, Vol. 3, pp. 2–40.

Shi Nai'an & Luo Guanzhong 1981. *Outlaws of the Marsh*. Translated by Sidney Shapiro. Vol. 2. Beijing: Foreign Languages Press.

Sidelnikov, Viktor 1964. *Kazakhskie narodnye skazki*. [Kazakh Folktales.] Vol. 3. Alma-Ata: Kazakhskoe gosudarstvennoe izdatel'stvo khudozhestvennoi literatury.

Siig, Kristo 2013. *Suure härja müüt läänemeresoome rahvaste traditsioonis ja selle ajalooline taust*. [Myth of the Big Bull in the Tradition of Balto-Finnic Peoples and Its Historical Background.] Bachelor's thesis. Tartu: Tartu Ülikooli filosoofia teaduskond.

Tambiev, Pago 1900. Adygskie teksty. [Adyghe Texts.] *Sbornik materialov dlia opisaniia mestnostei i plemen Kavkaza*, Vol. 27, Part 4, pp. 1–62.

Taube, Erika 1994. *Skazki i predaniia altaiskikh tuvintsev*. [Tales and Legends of the Altai Tuvinians.] Moscow: Vostochnaia literatura.

Tedre, Ülo 1969. *Eesti Rahvalaulud. Antoloogia*. [Estonian Folk Songs. Anthology.] Vol. 1, fasc. 1. Tallinn: Eesti Raamat.

Tugov, Vladimir 1985. *Abazinskie narodnye skazki*. [Abazin Folktales.] Moscow: Nauka.

Uther, Hans-Jörg 2004. *The Types of International Folktales: A Classification and Bibliography. Based on the System of Antti Aarne and Stith Thompson*. Parts 1–3. Folklore Fellows' Communications, Vol. 284–286. Helsinki: Academia Scientiarum Fennica.

Volkov, A. & Maiorov, I. 1959. *Karakalpakskie narodnye skazki*. [Karakalpak Folktales.] Nukus: Karakalpakskoe knizhnoe izdatel'stvo.

Yegorov, V. 1978. *Kalmytskie narodnye skazki*. [Kalmyk Folktales.] Elista: Kalmytskoe knizhnoe izdatel'stvo.

Yevseev, Viktor 1950. *Karel'skie epicheskie pesni*. [Karelian Epic Songs.] Moscow & Leningrad: Izdatel'stvo Akademii nauk SSSR.

Zamaletdinov, Lenar 2009. *Tatarskoe narodnoe tvorchestvo*. [Tatar Folklore.] Vol. 3. Bytovye skazki. [Realistic Tales.] Kazan: Magarif.

Zograf, Georgy 1976. *Skazki narodov Indii*. [Tales of the Peoples of India.] Leningrad: Khudozhestvennaia literatura.

http://dx.doi.org/10.7592/FEJF2016.64.colour

FROM LISTING DATA TO SEMANTIC MAPS: CROSS-LINGUISTIC COMMONALITIES IN COGNITIVE REPRESENTATION OF COLOUR

Mari Uusküla, David Bimler

Abstract: When a free-listing task is used to elicit verbal concepts from a given semantic domain, it provides two indicators of the salience of each word for that linguistic community. These are the proportion of the subjects who include a word in their lists, and its average ranking priority position across the lists. The data also contain cues about the cognitive representation of the semantic domain, and in particular about the conceptual closeness among words. Closely associated words tend to prime each other and to appear in the lists in close succession. Clusters of mutually associated terms can be recognised, listed in one another's company, although with different priority for different subjects. We applied this approach to the domain of colour terms, converting lists for fourteen European languages into matrices of inter-term similarity, for analysis with multidimensional scaling (MDS) and hierarchical clustering. Two-dimensional MDS solutions or 'maps' were typically required to reflect two competing criteria by which terms were sequenced. Speakers of each language tended to follow a salience gradient, but also made separate clusters of fully-chromatic concepts – colour terms in *strictu sensu* – and unsaturated or desaturated concepts defined primarily by lightness rather than by hue. This and other features recurred across the languages despite their geographical and phylogenetic diversity, as cross-cultural universals in colour language, in addition to the well-known regularities governing basic colour terms and the stages of colour-lexicon development.

Keywords: basic colour terms, Cognitive Salience Index, colour language, cross-linguistic comparison, linguistic typology, listing task, multidimensional scaling

INTRODUCTION

In the Method of Listing, a semantic domain is specified, such as 'colour terms' or 'animal names', and participants are asked to list all the examples of the domain they can think of, in the order in which they come to mind. These lists contain information about the salience of the terms (Borgatti 1999; Weller & Romney 1988). Thus the Listing Method is a useful tool for determining which colour terms in a given language deserve to be singled out as Basic Colour Terms or BCTs (Corbett & Davies 1997; Jrassati et al. 2012): that is, 'natural classes',

the default level of specificity for classifying and communicating about colours. Crucially, languages differ in the number of BCTs.

The concept was introduced by Berlin and Kay (1991 [1969]), although details of the definition and the criteria of 'basicness' have subsequently evolved (one might speak of a continuous scale of basicness underlying the dichotomy between 'basic' and 'non-basic' (Kerttula 2007)). The argument here is that a non-basic term features in the productive lexicons of fewer speakers than a BCT, and those who do list it, do so later in their sequences. Smith, Furbee, Maynard, Quick and Ross (1995) ranked English colour terms in order of decreasing salience, and separated them into three distinct classes: BCTs; non-basics, which they subdivided into 'opaque' and 'transparent'; and 'complex' terms (plus a 'residual' category, which includes textures and patterns).

Using the index *'i'* to identify the terms, we can write n_i as the number of speakers who listed the i-term, and n_i / N as the proportion, where N is the total number of sequences collected. We can also write mr_i as the mean rank of that term across those n_i sequences. Both indicators of salience have been published in a number of studies of a variety of languages (e.g. Davies & Corbett 1994a, 1994b; Davies & Corbett & Margalef 1995; Hippisley 2001). Urmas Sutrop (2001) combined them in the Cognitive Salience Index, $\mathrm{CSI}(i) = n_i / N / mr_i$. The goal of this study is to extend this approach.

A useful model for thinking about the listing task is the image of a 'semantic network' or a 'semantic map'. Terms are imagined as nodes in the network, connected by links, with activation spreading between the nodes which is faster along stronger and more direct links (e.g. Collins & Loftus 1975; Goñi et al. 2011). The participant initially accesses the network through some especially prototypical or culturally-salient node (e.g. 'red'), and then explores it more-or-less systematically, as one term prompts the recollection of its neighbours; tending to progress down the gradient of salience, at each step following a link or association to whichever node has received most activation, until the boundary of the semantic domain is reached.

Thus a corpus of listing sequences contains additional structure about the pattern of associations and inter-relationships among the terms (non-basic colour terms as well as BCTs), to be extracted by more sophisticated analysis. If participants tend to list a pair of terms in close proximity – one term often immediately preceding or following the other – this is suggestive of a close conceptual link between them. In a precedent from the domain of animal names, Henley (1969) obtained a matrix of estimated inter-name link strength from listing data by averaging the 'adjacency' between each pair of terms (i.e. the absolute difference in the terms' sequence positions) across participants. Although each individual's personal network is unique, here we treat them as

approximations of a single cultural consensus. That is, the individual lists are conceived as coming from that consensus network (perturbed by random fluctuations), which we set out to reconstruct.

Despite the network metaphor, it is convenient to summarise the pattern of associations in terms of a spatial model. Here items are represented as points, arranged in a low-dimensional space so that proximity between each pair of terms reflects the average adjacency of corresponding terms. Another useful model is a hierarchical-clustering 'tree': a structure of successively forking branches, with terms at the 'leaves', so that all the leaves arising from one branch define a cluster (nested within larger clusters). Again, the distance between leaves – measured down one branch to a fork and up the other branch – reflects (dis-)similarity. Shepard (1974) applied multidimensional scaling or MDS to a matrix of animal-name adjacencies, and extracted a parsimonious, two-dimensional spatial map with interpretable axes. In a further replication of Henley's work, Storm (1980) elicited lists from subjects across six age bands, although the matrices of pairwise differences resisted reduction to a low-dimensional MDS solution.

The same analysis is also applicable to recall data in which participants attempt to recall terms from a list that was read to them (for instance, Friendly 1979, whose examples again include animal names). The focus here is on the clustering of items in the MDS or tree solution, i.e. how items were organised by memory into thematic 'chunks'. Having recalled one item from a 'chunk', the participant finds it easier to access others and exhaust the semantic cluster before continuing to another (e.g. Goñi et al. 2011). We expect to encounter this modular structure here: self-contained sub-lists of colour terms, which flock together, though appearing at different positions in different participants' lists.

Although we are interpreting list adjacency in terms of 'similarity', this is a conceptual, not a perceptual connection. Associational connection can stem from a contrast, or an antagonism, as much as from the number of features the two items share. Cultural associations and collocations contribute to their mutual priming. We sourced listing data from 14 languages, from three different language phyla – Indo-European, Uralic and Altaic – though all were European, and in cultural contact with their neighbours. We are interested in the extent of commonality across the MDS solutions. In other words, do the different language communities impose similar patterns of thematic and dimensional organization upon their colour lexicon?

One development in colour linguistics has been the recognition of common threads across cultures in terms of BCTs and the corresponding perceptual categories. Languages vary in the size of their colour lexicon, and the terms themselves may not be exact counterparts. A parsimonious inventory of focal

hues accounts for a large proportion of linguistic variability (i.e. the prototype hues, chosen as the best example of each category), with languages diverging primarily in how many of these foci they recognise as the nuclei of categories (Regier & Kay & Cook 2005). In Guatemala, Harkness (1973) observed that speakers of Mam (a Mayan language) partitioned colour space into fewer colour categories than their Spanish-speaking neighbours, but the foci of those categories each had a counterpart from the Spanish foci. The Namibian language Himba and the Papua-New-Guinea language Berinmo both recognise five BCTs, but *dumbu* in Himba is not quite equivalent to *wor* in Berinmo (Roberson et al. 2004): different boundaries in colour space distinguish these categories from neighbouring categories. Nevertheless, their foci are very similar.

Thus it is reasonable to look for recurring patterns in colour-lexicon semantic networks, albeit purely cognitive patterns, in contrast to the perceptual / conceptual universals noted by Berlin and Kay.

METHOD

Procedure

Data were collected in the course of a standard interview for establishing the BCTs in a language, in which participants also performed a colour-naming task subsequent to the listing task (Corbett & Davies 1997). The participants were blind to the topic as they were recruited to 'answer questions regarding their native language'. They were requested to 'Please name as many colours as you know' in their L1. Time allocated for listing and the length of the list were not limited. All elicitation data was gathered orally and written down (and/or recorded) by the experimenter, except for the Swedish and English groups who wrote their responses.

Languages and participants

The sample contains 14 languages belonging to the Indo-European, Uralic and Altaic families. Between them, these languages are spoken in a vast territory of Eurasia. The distribution of languages according to family is presented in Table 1, along with numbers of participants and other information.

Present Udmurt participants were divided between speakers of Northern and Southern Udmurt dialects, in which 'pink' is translated as *lemlet* and *ljölj* respectively (many of the present participants knew both terms and listed them consecutively). Southern dialects also possess the BCTs *kuren'* 'brown' and

Language	Family	No of subjects? (F/M)	Mean age (age range)	Location of interviews	Data collected by/ reference	Mean length of each participant's list (F, M)	Total no of unique terms
Italian		133 (73, 60)	39 (11 to 80)	Florence	MU (2014)	19.86 (21.07, 18.38)	310
Castilian Spanish		38 (20, 18)	43 (22 to 85)	Madrid	Kelly Parker	15.05 (16.96, 12.94)	97
Swedish		16 (14, 2)	22 (19 to 35)	Nybro (Sweden)	Ivar Jung	56.25 (57.43, 48.0)	362
English	Indo-	34 (20, 14)	44 (18 to 70)	Wellington (NZ)	DB	23.46 (26, 19.88)	
Lithuanian	European	67 (48, 19)	41 (15 to 80)	Vilnius and Rokiškis	Simona Pranaitytė	20.33 (20.29, 20.42)	175
Latvian		50 (31, 19)	33 (15 to 86)	Rīga, Daugavpils	Maksims Ivanovs	14.60 (15.29, 13.47)	109
Czech		52 (33, 19)	35 (15 to 70)	Brno, Prague	MU	20.65 (21.70, 18.84)	213
Russian		24 (14, 10)	39 (20 to 61)	St. Petersburg	Elena Ryabina	20.83 (24.29, 16.00)	146
Estonian 1		50 (31, 19)	42 (9 to 77)	Tallinn, Tartu and other locations in Estonia	Margaret Raidlepp, Triin Kalda	28.18 (32.90, 20.47)	386
Estonian 2		80 (53, 27)			Urmas Sutrop (2002)	18.93 (19.51, 17.78)	282
Estonian 1+2		130 (84, 46)				22.48 (24.45, 18.89)	600
Finnish		68 (42, 26)	40 (10 to 75)	Helsinki, Turku and near Tampere	MU	22.15 (23.31, 20.27)	342
Komi-Zyrian	Uralic	51 (37, 14)	49 (11 to 81)	Syktyvkar and other towns or villages	Elena Ryabina (2011)	12.00 (12.16, 11.57)	112
Udmurt		130 (81, 49)	43 (9 to 80)	Republics of Udmurtia, Tatarstan and Bashkortostan	Elena Ryabina (2011)	16.72 (16.90, 16.43)	280
Hungarian		125 (66, 59)	36 (9 to 82)	Budapest, Debrecen; other Hungarian cities and villages	MU	22.87 (21.35, 24.58)	459
Turkish	Altaic	56 (30, 26)	36 (14 to 79)	Antalya and Ankara	Kaidi Rätsep (2011)	17.46 (20.33, 14.15)	172

Table 1.

nap-čuž 'orange' but in Northern dialects no single term is assigned to orange and brown stimuli consistently enough to qualify for basicness. The language possesses the term *busir* 'purple', but this is not known to all language speakers (Ryabina 2011).

All subjects were recruited volunteers, with a representative sample of speakers – as a rule of thumb, Weller and Romney (1988) recommend a minimum of 20 to 30 informants. The sampling was random. Participants had different dialectal, educational and social backgrounds. The interviews were carried out by a native speaker (for Lithuanian, English, Estonian, and Udmurt) or a proficient L2 speaker (for all other languages). All subjects were screened for normal colour vision using pseudoisochromatic plates (Ishihara 2008) or the City University test (Fletcher 1980).

Data for 80 of the 130 Estonian subjects were collected by Urmas Sutrop in his 2002 study.

For three of the languages, additional data were available from participants in a separate study that focussed on terms used in the blue region of colour space (Bimler & Uusküla, in preparation). That study featured colour sorting as well as the naming task, but not all participants performed the listing task. Specifically: of the 133 Italian subjects, 102 participated in the main study, while 31 lists came from 54 participants in the second study. Of the 67 Lithuanian subjects, 51 participated in the main study, while 16 lists came from 50 participants in the separate study. Of the 130 Udmurt subjects, 125 participated in the main study, while 5 lists came from 25 participants in the second study.

Analysis

As noted above, the Cognitive Salience Index for items generated in listing data (Sutrop 2001) combines two sources of information: the proportion of participants to whom a term occurs, and the stage at which it occurred to those participants, i.e. its priority in their lists.

$$\mathrm{CSI}(i) = n_i / N / mr_i$$

Smith (1993) defined a salience index S, which is similarly a confluence of frequency and priority (also Smith et al. 1995). However, the tables of n_i and mr_i that have been published for a number of languages (e.g. Davies & Corbett 1994a, 1994b; Davies & Corbett & Margalef 1995; Hippisley 2001) are not sufficient for calculating S(i).

Following Henley (1969) and Friendly (1977), we defined a measure of adjacency ADJ(i,j) between the i-th and j-th terms. A matrix of separations SEP_p is obtained for the p-th subject, with entries

$$SEP_{pij} = \ln \mid s_{pi} - s_{pj} \mid .$$

This definition incorporates information from listing data beyond immediately-sequential items. Even if a participant does not list term B directly after term A, we assume that the activation of A persists, increasing the probability that B will be listed subsequently; with the corollary that more remote separation, such as the appearance of B two steps rather than three steps after A, is still an indication of the association between them.

The number of participants for whom SEP_{pij} is defined (because they included both terms i and j in their sequences) is c_{ij}, the co-occurrence of that pair of terms. Then ADJ(i,j) uses the mean of the separations, averaged across only those participants:

$$ADJ(i,j) = \exp [(\Sigma_p SEP_{pij}) / c_{ij}].$$

The effect of the logarithmic transform in SEP_{pij} is that ADJ is the *geometric* mean of absolute differences rather than the *arithmetic* mean. This is so that an increment to a small separation (such as the difference between two consecutive terms, where $SEP_{pij} = 0$, and terms separated by an intermediary, where $SEP_{pij} = \ln 2$) has more effect on the value of ADJ(i,j) than the same increment for more widely separated terms.

Not all the terms nominated in each language were included in the analysis; only the more salient ones, for which the associations ADJ(i,j) with other terms could be estimated with some confidence. The reliability of each estimate depends on the co-occurrence c_{ij}, and is limited by statistical fluctuations: ADJ(i,j) is less reliable when the i-th and j-th terms are both low-salience so they co-occur only in the lists of a few unrepresentative participants. Naturally this threshold of confidence allowed more terms to be considered in language samples for which more participants were recruited, or if these recruits were relatively productive.

Adjacency matrices were analysed with MDS, using the PROXSCAL software within SPSS to represent terms as points in a low-dimensional space. Ordinal transforms were used (the non-metric form of MDS), with the tie-breaking option. We retained two-dimensional solutions to allow them to be compared across languages, including those for which fewer items could be analysed reliably and a third axis could not be justified.

PROXSCAL allows the entries in a proximity matrix to be weighted by reliability (less-reliable entries having less influence in determining the locations of

the two corresponding items). We entered the square root of co-occurrences, $\sqrt{c_{ij}}$, for this purpose. To ensure that the algorithm converged towards the overall best fit rather than being trapped in a 'local minimum', starting configurations were randomised, iterating over 500 random starts for each matrix.

Hierarchical clustering was applied using Ward's algorithm, implemented in SPSS 20. Note that non-metric models are not applicable for HCA, i.e. the analysis necessarily assumes that the (postulated) underlying similarities among items follow a linear relationship to the estimates obtained from the data. In addition, there is no option for including information about the reliability of individual estimates.

RESULTS

Some basic statistics for the lists are included in Table 1: mean length of each participant's list (with separate values for males and females), plus the total number of unique terms nominated. Productivity ranged from a mean of 12 terms per list for Komi-Zyrian, up to 56.25 for Swedish. Swedish was a special case; the instructions were evidently understood differently than in other samples, as if participants felt called upon to list *all* the terms they could think of, perhaps to make up for the small sample size. Swedish participants were all art or design students, who recorded their lists in writing, in constrast to other groups who gave oral responses. The distributions of terms per list $list_p$ were strongly skewed, so the following analyses were performed on $\log(list_p)$ to produce more normal distributions.

ANOVA confirmed the reality of this variation, with a significant effect from the independent variable 'language group' ($F = 14.25$, 13 d.f., $p < 0.001$). The contribution to variance from the variable 'gender' was also significant, with women tending to list more terms than men ($F = 21.33$, 1 d.f., $p < 0.001$). However, this was not consistent across languages, i.e. a gender-by-language interaction term also reached significance ($F = 2.24$, 13 d.f., $p = 0.07$). In analyses of each sample in isolation, the male/female difference was significant for five languages: English: ($t = 2.43$, $p = 0.02$), Estonian ($t = 2.85$, $p = 0.005$), Italian ($t = 2.13$, $p = 0.035$, Spanish ($t = 5.70$, $p = 0.022$), and Turkish ($t = 3.21$, $p = 0.002$).

Figure 1(a) shows CSI values for the 25 most salient terms in Czech, while Figure 2(a) is the MDS solution for the most frequent terms, using Czech as a simple but representative case. When terms are ranked in order of decreasing salience, the values typically follow a roughly exponential decline. Figure 1(a) exploits this generalisation by plotting CSI on a logarithmic scale, to render the decline roughly linear.

Figure 1. *Ln(CSI) for most salient terms (ranked in order of decreasing salience) for (a) Czech; (b) English; (c) Estonian; (d) Finnish; (e) Hungarian; (f) Italian; (g) Komi-Zyrian; (h) Latvian; (i) Lithuanian; (j) Russian; (k) Spanish; (l) Swedish; (m) Turkish; (n) Udmurt.*

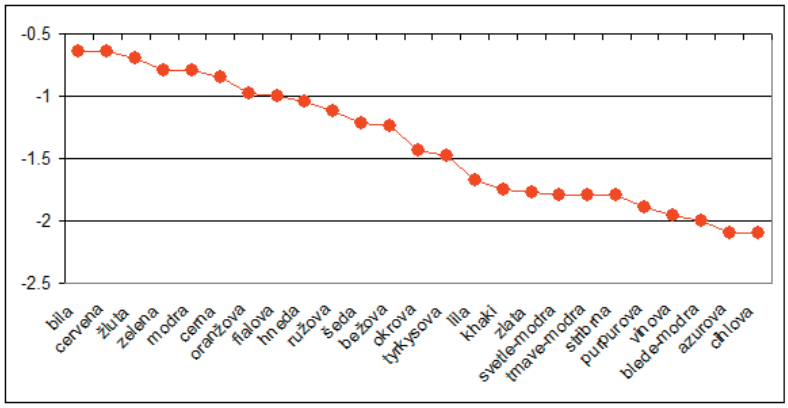

Figure 1(a). *Czech.*

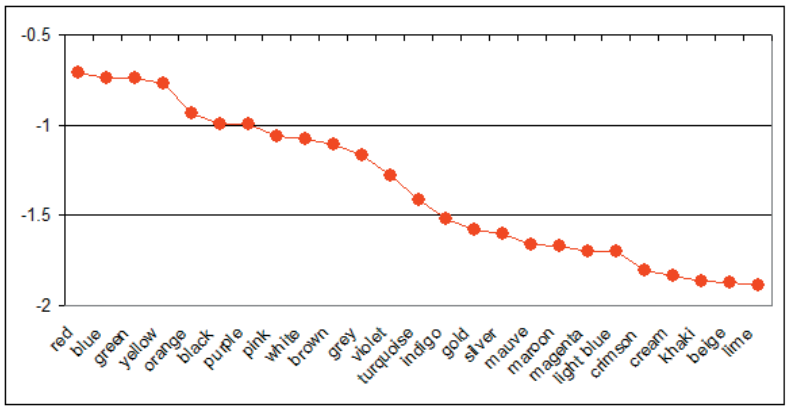

Figure 1(b). *English.*

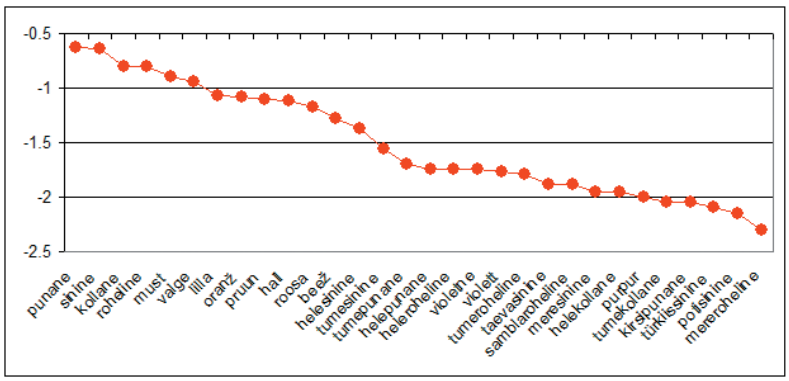

Figure 1(c). *Estonian.*

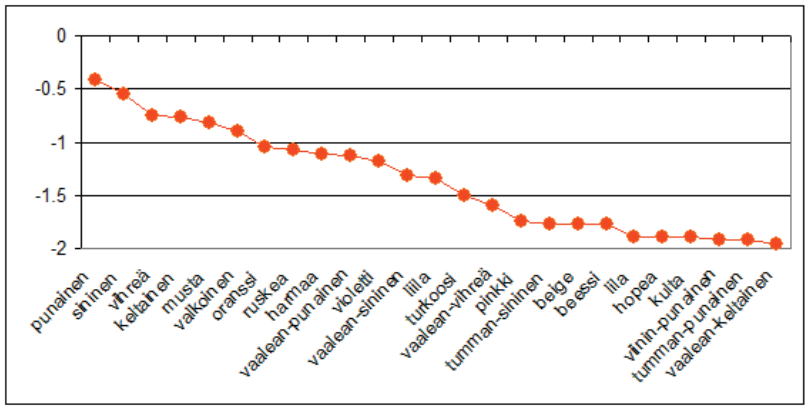

Figure 1(d). *Finnish.*

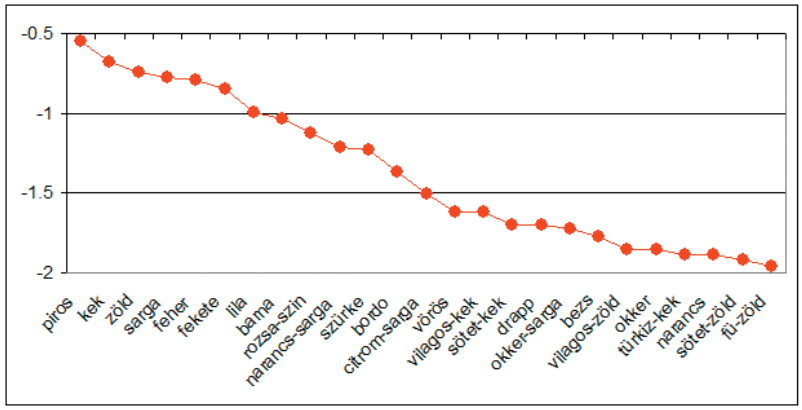

Figure 1(e). *Hungarian.*

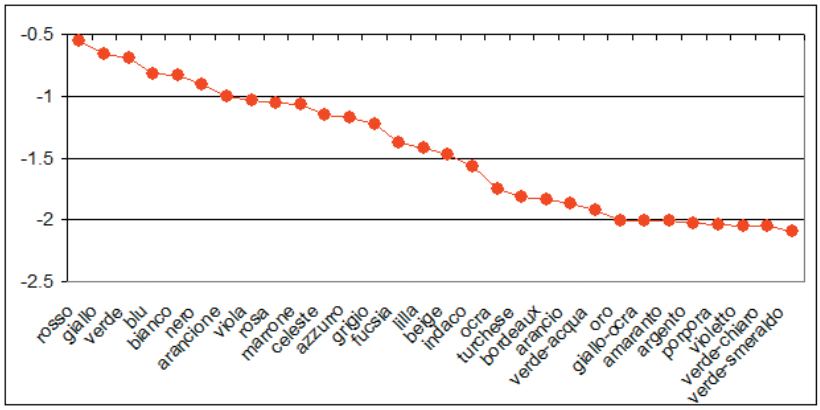

Figure 1(f). *Italian.*

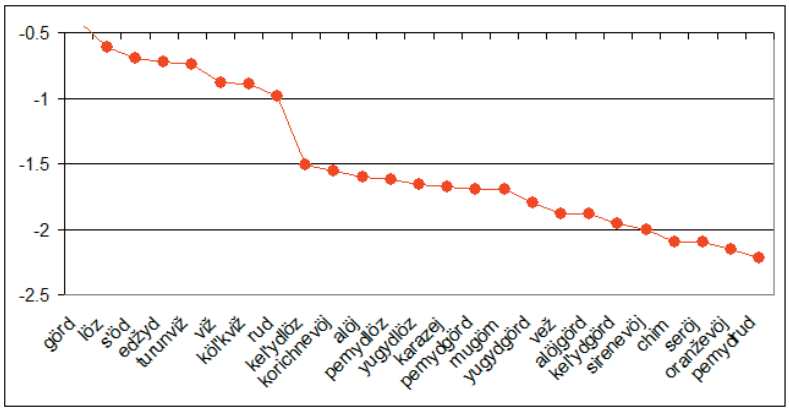

Figure 1(g). *Komi-Zyrian.*

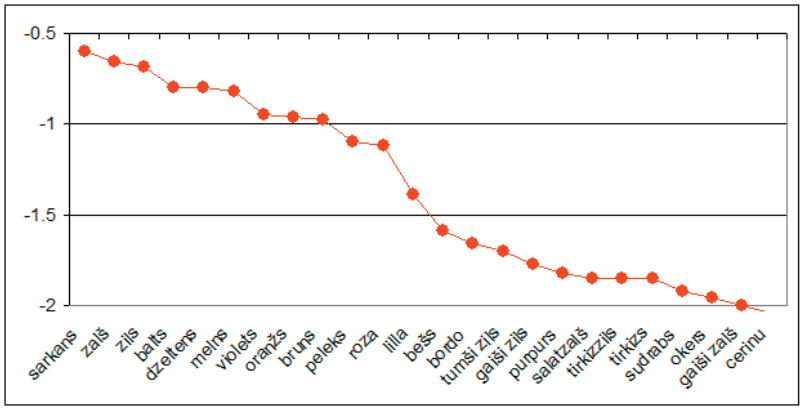

Figure 1(h). *Latvian.*

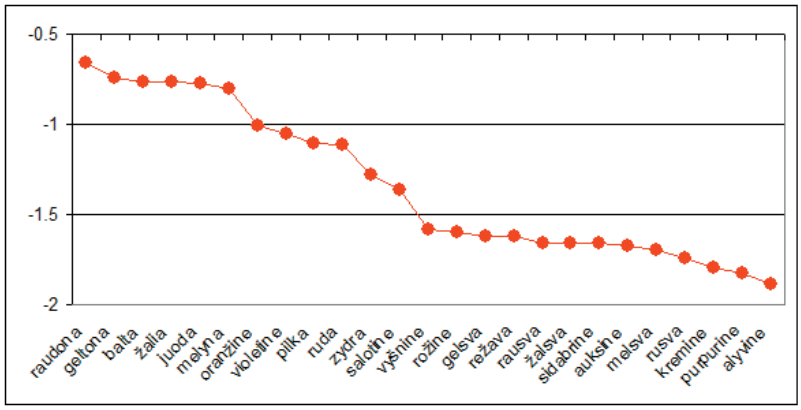

Figure 1(i). *Lithuanian.*

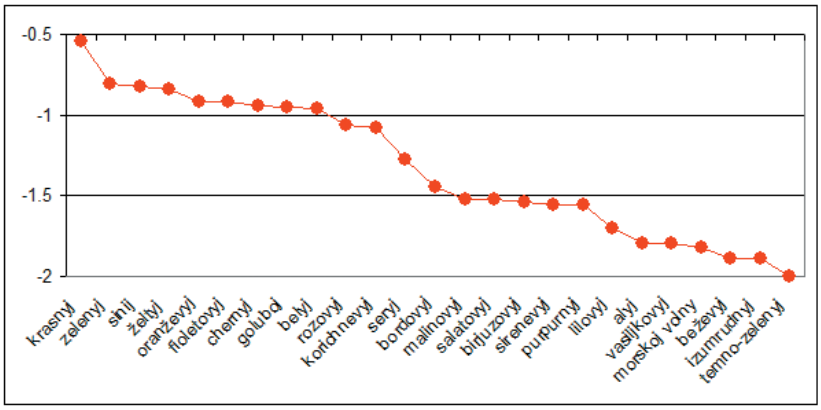

Figure 1(j). *Russian.*

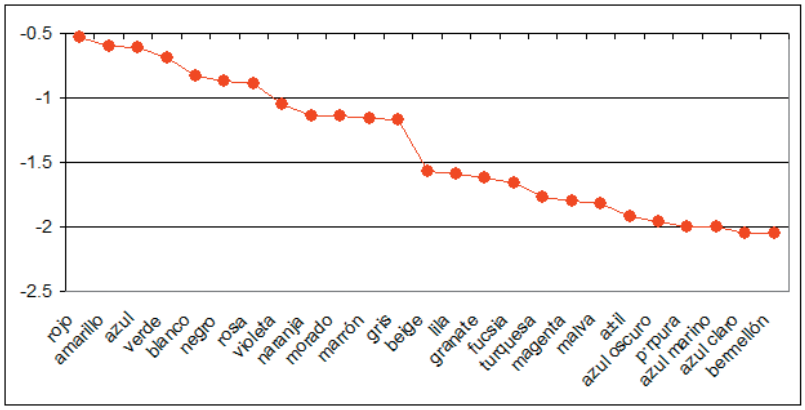

Figure 1(k). *Spanish.*

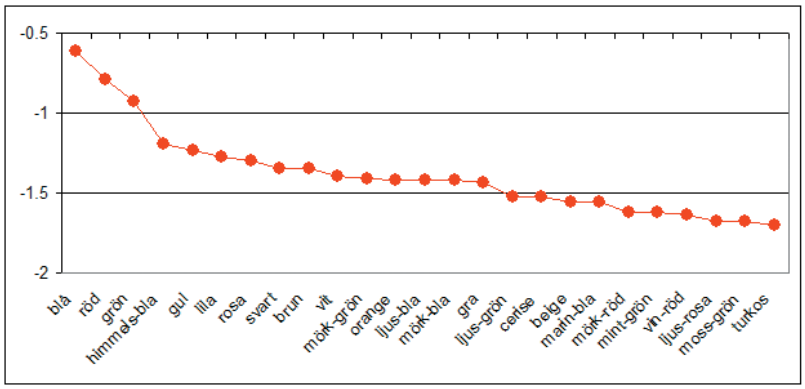

Figure 1(l). *Swedish.*

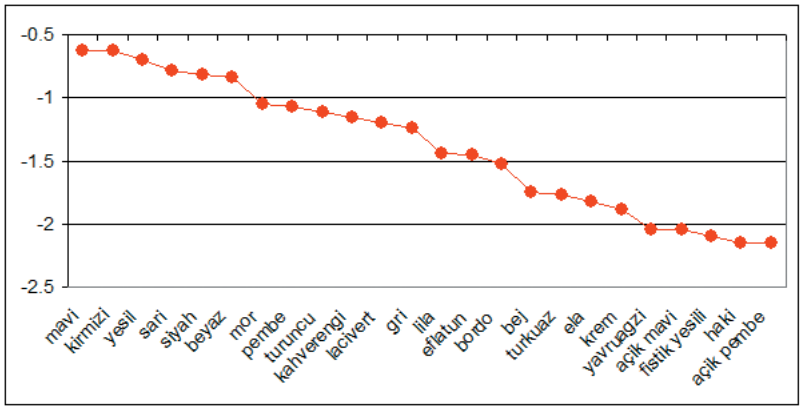

Figure 1(m). *Turkish.*

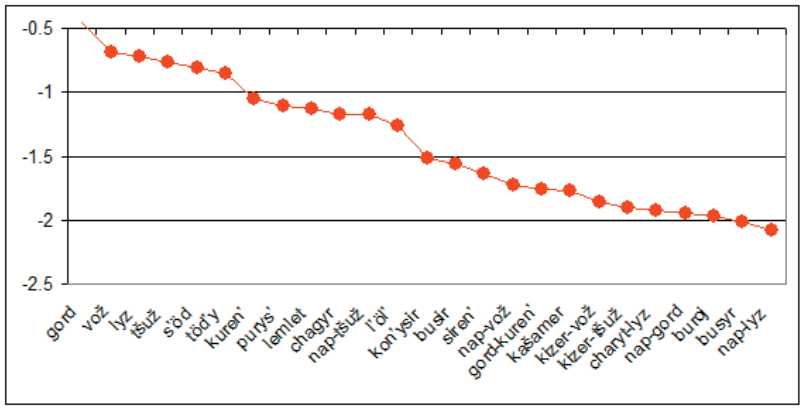

Figure 1(n). *Udmurt.*

In the MDS plot, the horizontal axis D1 can be identified as a gradient of salience. It ranges from primary terms at the left (starting with *červená* 'red') up to non-basic terms at the right, including the marginal cases of *stříbrná* 'silver' and *zlatá* 'gold'. Between these extremes lie terms that are basic but secondary.

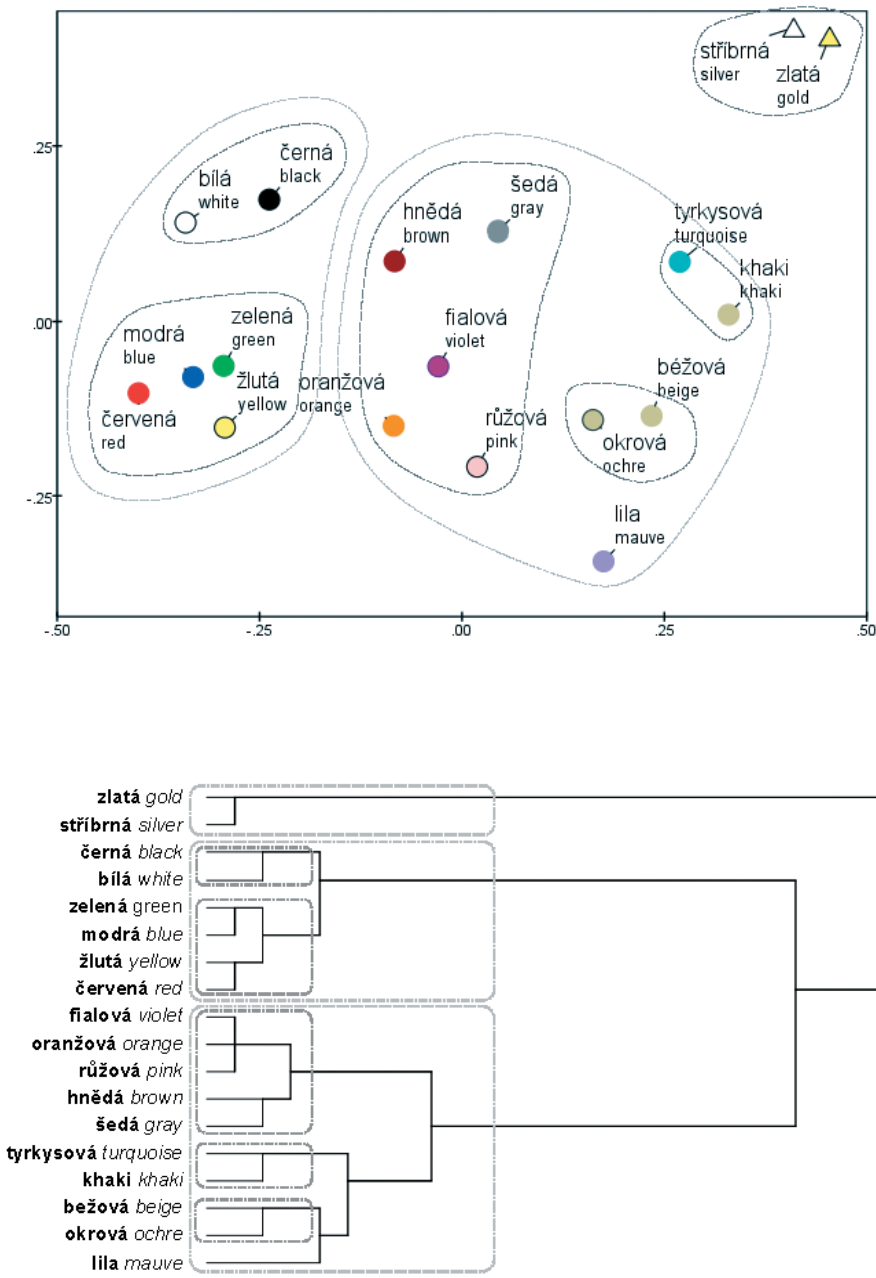

Figure 2. *Inter-term adjacencies within Czech data represented as (a) MDS solution; (b) dendrogram.*

If everyone listed terms in the same order, a single axis would suffice, but this was not the case; additional axes are required to accommodate the variations among participants. Within the *primary* basic terms, the vertical axis D2 reflects a distinction between 'cardinal hues' and the achromatic primaries *černá* 'black' and *bílá* 'white', which some people list before the cardinal hues while others list them later. A parallel distinction prevails among the secondary terms, separating *oranžová* 'orange' and *fiolová* 'purple' on one hand – defined by their chromatic content – from *hnědá* 'brown' and *šedá* 'grey' on the other hand, for which desaturation or lightness are the characteristic qualities (with *růžová* 'pink' conceptualised as part of the former chromatic group). At the right-hand extreme of D1, D2 still serves to distinguish *stříbrná* (silver) and *zlatá* (gold) – hardly colours at all – from other non-basics.

To put it another way, if one is listing primary terms and has reached 'black' or 'white', it is easy to segue to the achromatic secondary terms. Conversely, from the cardinal-hue primaries it is easier to continue the theme of chromatic content by jumping to secondaries like 'orange' and 'purple'. That is, the terms at each level of salience are organised in parallel fashion. Pairwise similarities *between* salience levels are not all the same, but reflect (or are best represented by) global organisational attributes (or in the spatial model, dimensions).

The outcome of hierarchical clustering for the Czech data is plotted in Figure 2(b), as a tree diagram (dendrogram). Vertical lines slice through the branches at two levels to emphasise the patterns of clusters and singletons remaining at those stages of agglomerative clustering. The same clusters are superimposed upon the spatial model (Figure 2(a)) by enclosing the clustered items within loops. These clusters highlight the 'chunking' of terms in semantic memory, where the items comprising a chunk tend to emerge as self-contained sub-lists within the listing sequence. However, clustering cannot show the parallelism of internal structure and the relationships between chunks. We do not show the full dendrograms for the remaining languages, only the intermediate-level clusters, again superimposing them on MDS solutions.

Figure 1 includes salience values for the other languages. As noted above, when terms are ranked by salience, they follow a roughly exponential decline. This linear decline of ln(CSI) serves to emphasise any departures from the exponential trend, and allows a second generalisation: that the decline is punctuated by sudden drops separating stretches of slower decline. One especially prominent step distinguishes the BCTs from the various hyponyms and complex terms used for finer chromatic distinctions or as lexical ornaments. This is most clear in Latvian, Lithuanian and Spanish and least perceptible in the cases of Swedish and Hungarian. In many cases, a smaller step can be discerned separating the 'primary' BCTs (the four 'cardinal hues' plus the antithetical pair of Black and White) from the 'secondary' BCTs – an inventory of some or

all of the counterparts of English orange, purple, brown, grey, pink and light blue. The Swedish plot is anomalous, suggesting that more than 16 lists are ideal for the results to stabilise.

Salience on its own is sometimes misleading as an indicator of basicness. Spanish *violeta* 'violet' and *morado* 'purple' are both more salient than *gris* 'grey', while in Czech *bežová* 'beige' falls on the BCT side of the step. The Estonian equivalent *beež* enjoys a similar status. Conversely, Russian *seryj* 'grey' slips down to the non-basic side of the step, as do Lithuanian *rožine* and *ružava* 'pink' which are both less salient than *vyšninė* 'cherry'.

In Komi-Zyrian the initial gradient of putative BCTs extends for eight terms, the eighth being *rud* 'grey', followed by a step to the modified 'blue' forms *kelyd'löz* and *pemydlöz*. Subsequently *alöj* 'pink' and *koričnevöj* 'brown' appear, neither evidently salient (the latter is a loan-word from Russian); *oranževöj* 'orange' is 25th in order of salience and not in wide use.

In Udmurt the BCT gradient extends for 11 terms. It includes *lemlet* 'pink', but *ljölj* – its counterpart in Southern dialects – falls along the steeper decline to the non-basic segment of the plot. The next non-basic term is *busir'* 'purple'. The ninth and tenth basic terms are *nap-čuž* 'orange' and *purys'* 'grey', while the 11th is the ostensibly basic *čagyr* 'light blue'.

Languages like Udmurt, in which more than one BCT share the region of colour space spanned by the English BCT 'blue', are of special interest because of the challenge they pose to the dogma that the ceiling of complexity is set at 11 BCTs. The canonical case is Russian, where multiple lines of inquiry converge in supporting the basicness of *goluboj* 'light blue' (Paramei 2005, 2007). In Figure 1(j) it is the eighth most salient term. Lithuanian is a second candidate: like Udmurt, there has been linguistic pressure from neighbouring Russian-speaking populations, perhaps contributing to the basicness of the 'light-blue' term *žydra*. Figure 1(i) ranks *žydra* as the 11th most salient term, at the beginning of the step from BCTs down to non-basic terms, ahead of *salotinė* 'light green'. In Italian the two light-blue terms *azzurro* and *celeste* are the 11th and 12th most salient terms, ahead of *grigio* 'grey', which is followed by an abrupt drop in salience to *fucsia* 'fuchsia'. In contrast, in Turkish it is the dark-blue term *lacivert* that may be basic (Özgen & Davies 1998), complementing the broader 'blue' term *mavi*. The status of *lacivert* remains moot (Rätsep 2011) but it is the 11th most salient term in Figure 1(n), again ahead of 'grey' *gri*, attesting to its.

Turning now to the remaining MDS solutions, plotted as Figures 3 to 15, many of the features of Czech recur. We noted that *hnědá* 'brown' appears near *šedá* 'grey', with a positive D2 value, as if conceptualised by its darkness or desaturation. This is generally true, with the exception of Spanish (where *marrón* is grouped with chromatic terms including *granate* and *fucsia*).

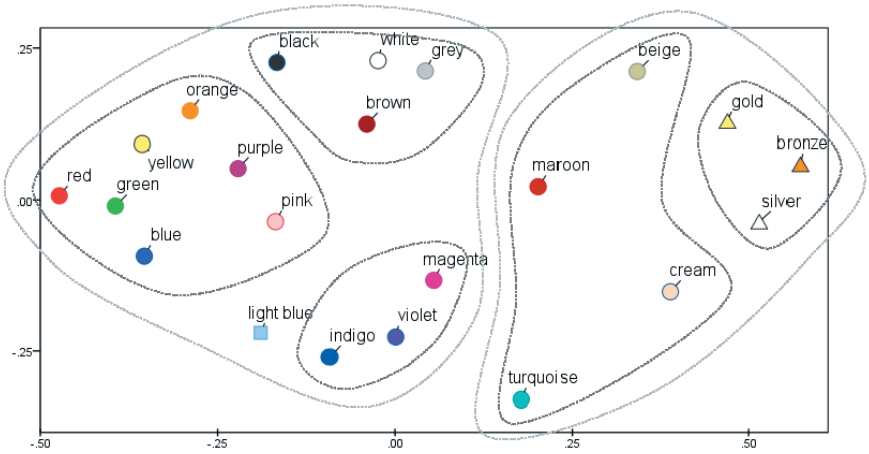

Figure 3. *English.*

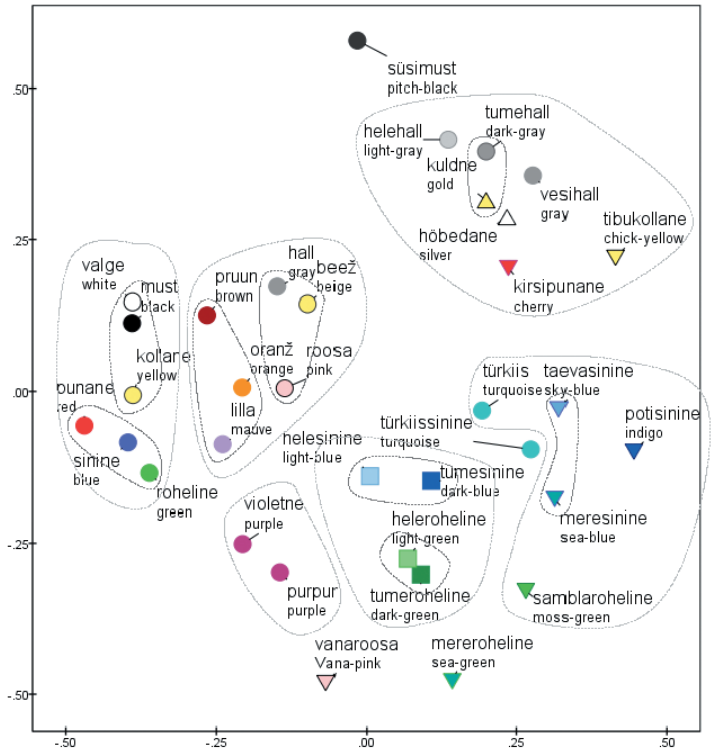

Figure 4. *Estonians.*

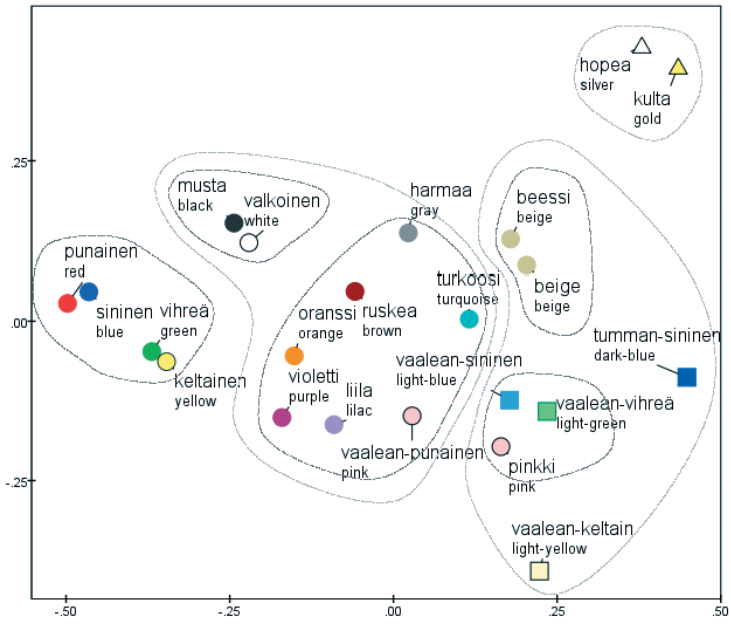

Figure 5. *Finnish.*

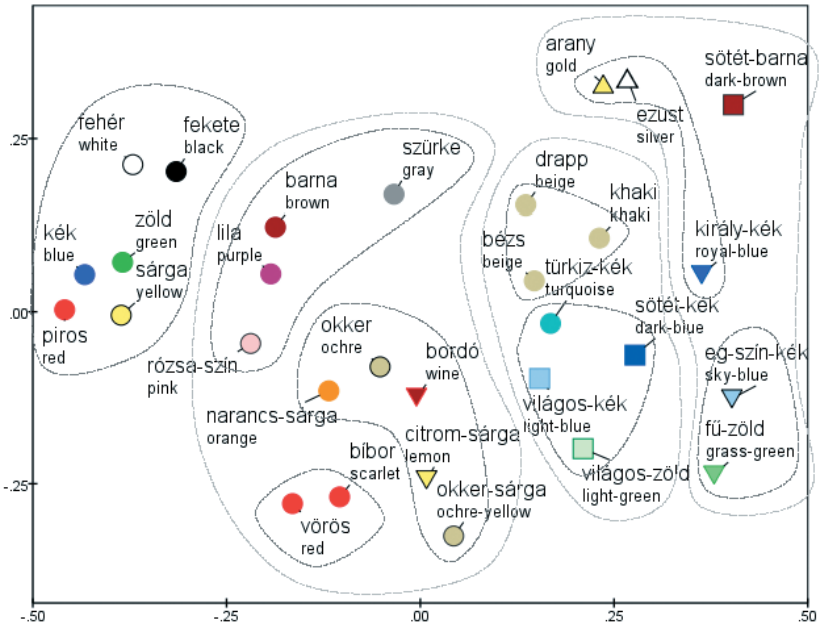

Figure 6. *Hungarian.*

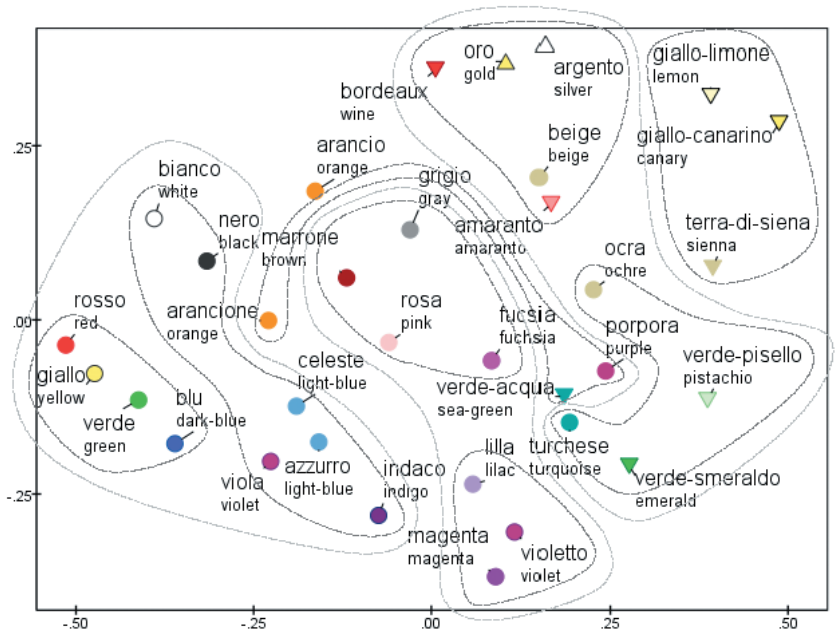

Figure 7. *Italian.*

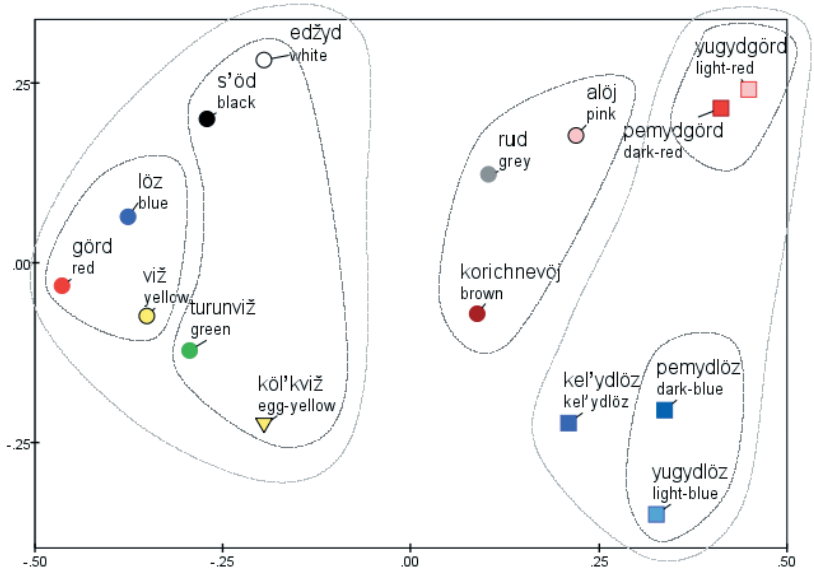

Figure 8. *Komi.*

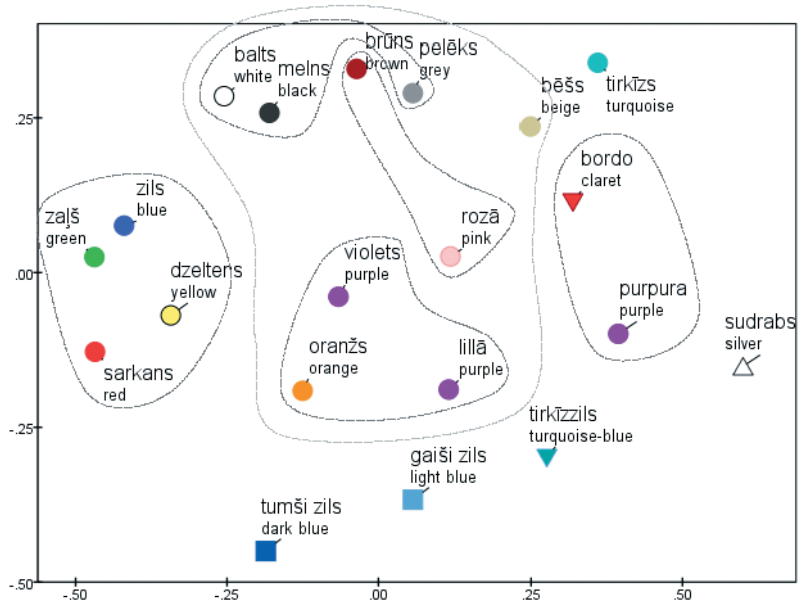

Figure 9. *Latvian.*

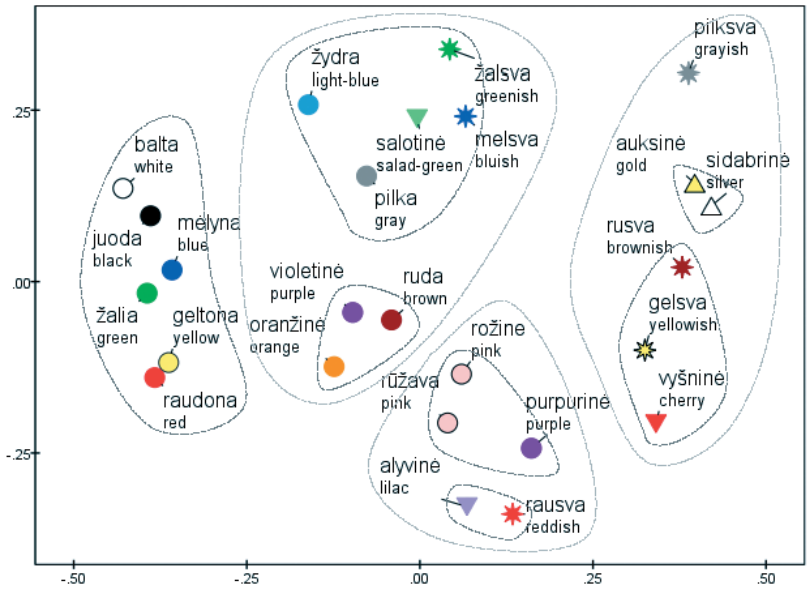

Figure 10. *Lithuanian.*

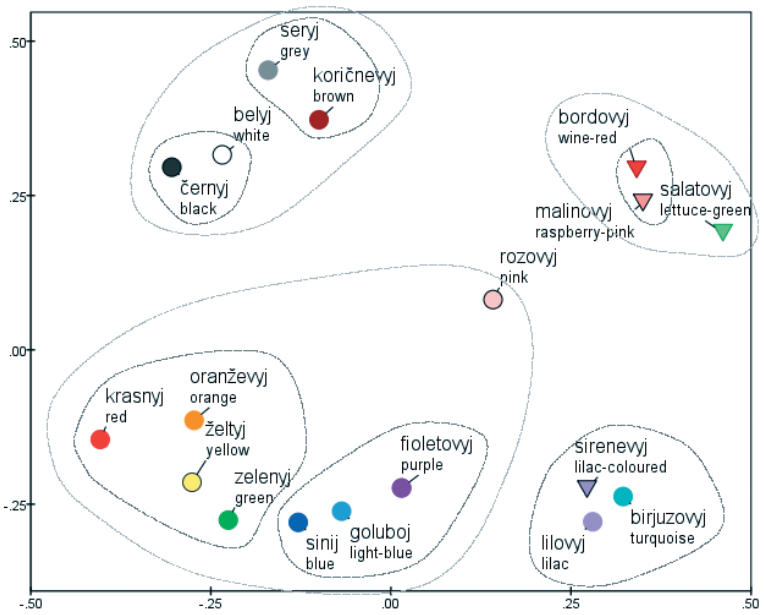

Figure 11. *Russian.*

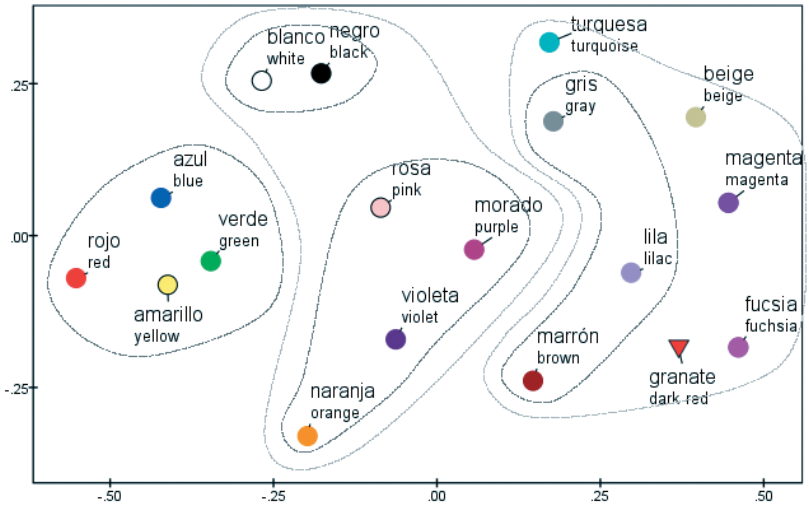

Figure 12. *Spanish.*

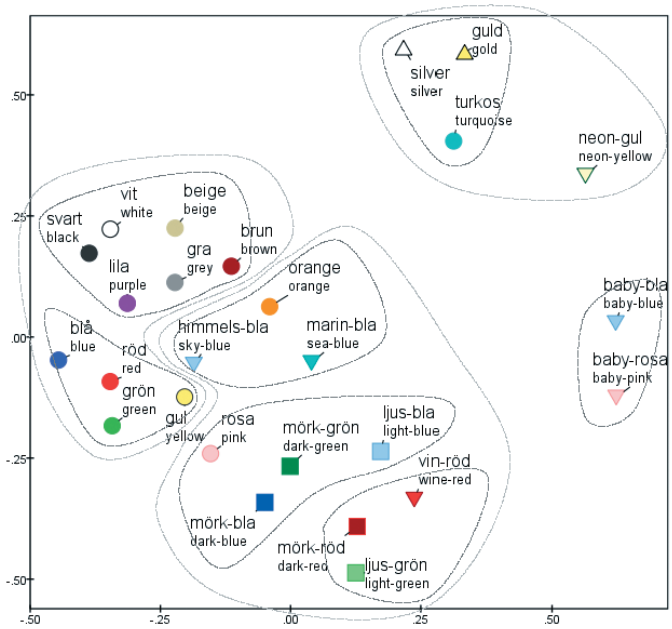

Figure 13. *Swedish.*

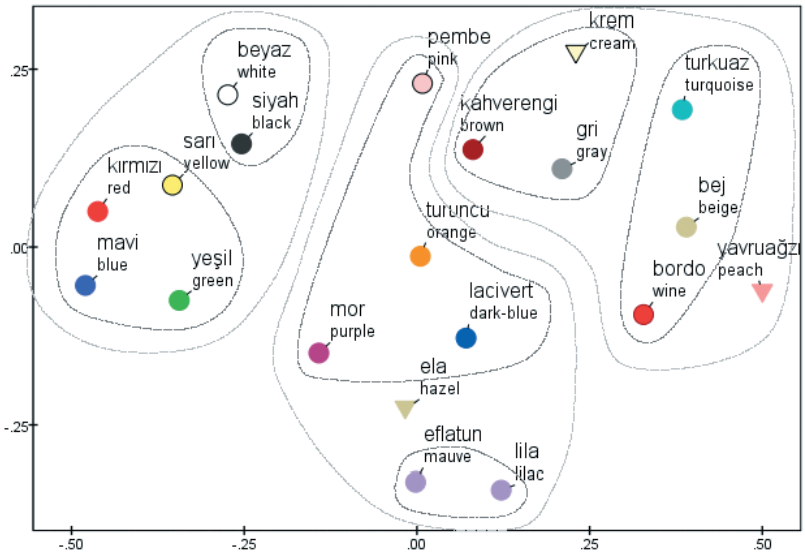

Figure 14. *Turkish.*

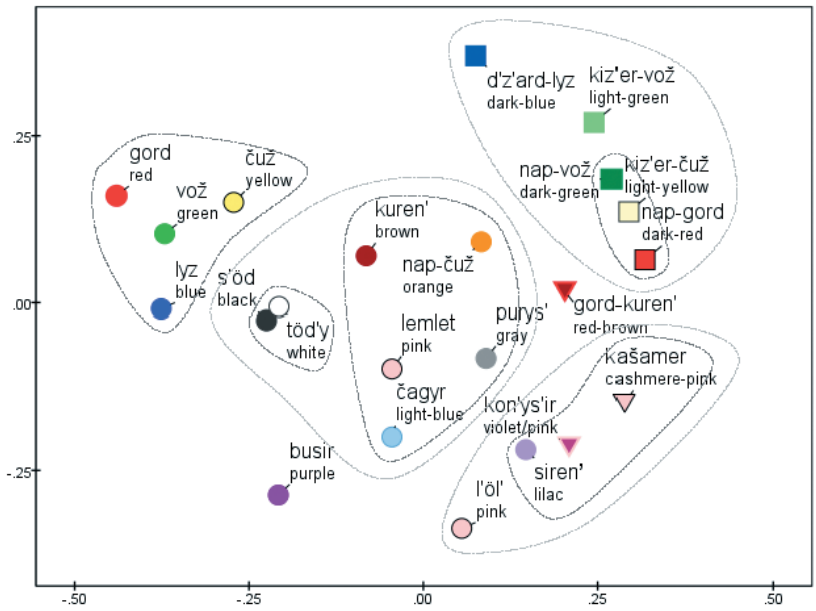

Figure 15. *Udmurt.*

In another generalisation, the equivalents of 'pink' were listed among and gravitated towards the secondary terms 'purple' and 'orange', as if regarded as a fully chromatic concept. Turkish is an exception. *Pembe* 'pink' was often listed in the company of achromatic terms (including *kahverengi* 'brown' and *krem* 'cream') placing it high on D2, as if linked to them for Turkish participants by its connotations of lightness and desaturation.

In the Udmurt solution, both variants of 'pink' (*ljölj* and *lemlet*) lie at one extreme of D2, near *busir* 'purple'. As noted, however, *busir* is not universally salient. The other extreme of the second dimension of the Udmurt MDS solution is dominated by a clump of qualified terms (blues and greens), while *purys'* 'grey' and *kuren* 'brown' are located centrally. A straightforward interpretation of this axis is not satisfactory.

Also of note are terms translatable as 'lilac' or 'mauve', in languages where they were sufficiently frequent (e.g. *lila* 'lilac' and *eflatun* 'mauve' in Turkish, *lila* in Finnish). Again these terms were associated with and listed with 'purple' and other 'true colours', rather than with the achromatic or marginal colour terms.

The larger data sets for some languages allow the analysis to include more non-BCTs, generally forming a tier of low salience at the right-hand extreme of

D1 in the MDS solutions. Note that D1 is best interpreted as the 'priority' aspect of salience, rather than as 'frequency'. A term can be relatively infrequently used, but still lie to the left of D1, if those people who do list it, do so early in their lists, in the context of high-salience terms. This is most apparent for terms for which alternative forms or dialect variants exist, with participants listing only one or the other form, but in similar contexts – resulting in adjacent points in the MDS solution. Thus the Lithuanian participants are split between *rožine* and *ružava* 'pink', reducing the frequency of both forms, but both are high-priority (it is possible that neither word is basic; see Pranaitytė 2011). Other examples are *beige* and *beessi* 'beige' in Finnish, *türkiis* and *türkiissinine* 'turquoise' in Estonian, and in Italian *arancione* and *arancio* 'orange', where less than a sixth of participants used the latter truncated form. The Finnish BCT counterpart to 'pink' is *vaalean-punainen*, literally 'light red', although conceptually understood as a whole (Uusküla 2007). *Pinkki* was only listed by 16 of 68 participants, but those 16 tended to list it just before or after *vaalean-punainen*, testifying to its near-synonymity and providing it with a high priority (for finer distinctions within 'pink' in Germanic languages, see Vejdemo et al. 2015).

Returning to less-salient terms, in the Uralic languages these tend to be 'transparent' terms coined by combining BCTs (e.g. 'red-brown'), or by modifying them with a qualifier of lightness or saturation ('light green', 'dark blue') or a metaphorical reference (e.g. 'sea-green'), including the dyeing process in the case of Estonian *potisinine* 'indigo', literally 'pot-blue'. The modified terms coalesce in distinct domains within the MDS maps, each domain grouped by a common source BCT (rather than by a common modifier). Typically in the data these terms are listed in chunks: a systematic attempt to exhaust all variants of (say) 'blue' before moving on to (say) 'green'. In addition, the MDS maps suggest that the sequence of these sub-lists tends to echo the earlier sequence of the BCTs themselves, for the arrangement of these domains is similar (though on a larger scale) to the arrangement of the BCTs in the 'basic' region of the map (for instance, *tume-hall* 'dark grey' and other modified greys in the Estonian map receive high D2 coordinates that locate them close to *kuldne* 'gold', echoing the role of D2 in distinguishing *hall* 'grey' itself from chromatic secondary BCTs). Typically 'green' and 'blue' are particularly generative, and the derived forms are characterised by their chromatic content, judging from their extreme positions on the D2 axis (see Figures 4, 6, 8; also Latvian and Swedish, Figures 9 and 13).

Figure 10 for Lithuanian displays a similar peripheral fringe of 'nuance' terms such as *žalsva* 'greenish'. The morphology of Lithuanian promotes the formation and the acceptance of such terms (Pranaitytė 2011). In contrast, in English hedged terms of the form 'X-ish' are valid descriptions for non-prototypal examples of a colour, but they are seldom regarded as colours *per se*, and Smith, Furbee, Maynard, Quick and Ross (1995) reported the form to be rare

across the 353 lists they elicited. *Žalsva* and *melsva* 'bluish' were linked with desaturated concepts.

This leads us to terms cognate with 'turquoise', which were salient in many languages. They reveal an interesting dichotomy in the way it is conceptualised. In Estonian, *türkiis* 'turquoise' and *türkiis-sinine* 'turquoise-blue' both belong to the cluster of qualified blues. This also occurs for Hungarian *türkiz-kék*. In Italian, *turchese* is located within a sector of qualified greens with *verde acqua* 'sea-green' and *verde smeraldo* 'emerald green' as its closest neighbours (while Russian *birjuzovyj* is clustered with a pair of 'lilac' descriptors). However, the Latvian data treated *tirkīzzils* 'turquoise blue' as a different colour from *tirkīz* 'turquoise': the former, regarded as a qualified 'blue', is located near *gaiši zils* 'dark blue', while the latter is at the opposite extreme of D2, near *bēšs* 'beige'. Similarly, Turkish *turkuaz*, Spanish *turquesa*, Swedish *turkos*, Finnish *turkoosi* and Czech *tyrkysová* are in the neighbourhood of 'beige' or 'grey' or the metallic sheens.

Equivalents of 'beige' were surprisingly salient in the listing data for several languages, e.g. *bész* (Hungarian) and *bej* (Turkish). Hungarian also possesses *drapp*, with a similar denotation as *bész* (Eessalu & Uusküla 2013). Participants seemed to focus on the non-chromatic connotations of the concept, associating and listing these terms with 'grey', or with 'gold' and 'silver' when the metallic sheens were listed.

Finally we return to the 'secondary blue' terms that complement the 'primary blue' in some languages. For Turkish, the MDS solution shows *mavi* 'blue' in the usual group of 'cardinal hues', while *lacivert* 'dark blue' was listed with the other chromatic secondary terms *mor* 'purple' and *turuncu* 'orange'. In Italian, a similar relationship emerged between the more inclusive primary term *blu*, and *azzurro* and *celeste* 'light blue', both mapped between *viola* 'violet' and *rosa* 'pink'. Of those two, *celeste* has stronger connotations of lightness and desaturation, and it was slightly closer to grey. The Udmurt term *čagyr* 'light blue' behaves similarly. In contrast, *žydra* in Lithuanian appeared as a desaturated concept in the data, associated with *pilka* 'grey', *salotinė* 'light green', and the nuanced terms *žalsva* and *melsva*, 'greenish' and 'bluish'.

Turning to Russian listings, the present results are aberrant in several respects. Data were sparse in this sample, with 24 participants recruited. Five of those 24 treated the cardinal hues and chromatic secondary terms together as a distinct 'chunk', listing them in a recurring rainbow sequence from *krasnyj* 'red' to *fioletovyj* 'purple'. The same pattern emerges in listing data from L1 Russian speakers in Estonia (Rätsep, pers. comm.). Presumably this reflects Russian pedagogy, which instils the colours of the rainbow with a suitable mnemonic, much as English-speaking children learn colour terms in a rainbow sequence with the ROYGBIV acronym (Paramei, pers. comm.). The consensus

across these five participants prevailed over the less-structured responses from the other 19 and distorted the location of terms in the MDS solution to accommodate a distinct band of rainbow terms, separated from other clusters of hues by large gaps (the mnemonic lists *goluboj* between *zelenyj* 'green' and *sinij*, but in the map its neighbours are *sinij* and *fioletovyj*).

DISCUSSION

Comparisons can be drawn with the semantic domain of animal names as a precedent, where the analysis of listing data to obtain a spatial map is long-established. The first dimension of such maps is often a gradient of familiarity or typicality, compatible with the frequency ranking reported by Henley (1969) – which started with 'dog, lion, cat, horse, tiger' and proceeded to less over-learned animals – or by Borgatti (1999) which began 'cat, dog, elephant, zebra…' A second dimension often makes a conceptual distinction between domestic and wild animals – the former category encompassing pets and farm animals, and the latter encompassing pests as well as 'zoo animals' (Shepard 1974).

Paulsen, Romero, Chan, Davis, Heaton and Jeste (1996) pooled the data of schizophrenic subjects to derive a map of animal name semantics, and drew conclusions from the differences between that map and one derived from normal controls (see also Aloia et al. 1996). That is, instead of assuming that all English speakers share a consensus animal-name semantic network – the assumption used here for colour terms – they allowed for the possibility of more than one consensus. The homogeneity of a group of subjects is a question of the agreement among their adjacency matrices $\mathbf{SEP}_p$. After reducing individual listings to adjacencies, the range of variation among them can be studied with factor analysis, although that line of inquiry is not pursued here.

In a special case of individual variation, it has been reported that women perform better than men at colour-related cognitive and perceptual tasks. In particular, they access a larger vocabulary when describing hue samples (Mylonas & Paramei & MacDonald 2014); these reports are not restricted to English-speaking samples (Rätsep 2013; Ryabina 2009). We therefore expected to find a gender difference in the number of terms each participant listed. The difference was indeed significant overall, but interestingly it varied significantly between samples, reaching the threshold of significance for only five languages in isolation, with some samples showing no difference at all. It is conceivable that men and women also differ in their patterns of listing, accessing different male and female consensus semantic networks. That possibility lies outside the scope of this study.

The list-based derivation of dissimilarity or adjacency used here is not the only possibility. It is essentially the same as Friendly's (1977) definition, except that ADJ(i,j) is the geometric rather than the arithmetic mean of list-wise ranking differences, resulting from the log transform of those ranking differences. This transform ensures a degree of diminishing returns in the definition: the presence of a term between terms A and B in a list indicates a higher dissimilarity than if A and B were consecutive; the presence of *two* interposed terms shows the dissimilarity to be higher again, but the *increment* in dissimilarity is smaller.

Borgatti (1999) used the list co-occurrences c_{ij} as a simple estimate of inter-item proximity, unweighted by the relative position of the i-th and j-th terms in each list. This is equivalent to assuming that once a term has been listed and the node representing it in the semantic network has been activated, it remains activated throughout the list, continuing to prime the nodes to which it is linked and raising the likelihood that they will also be listed. This analysis is dominated, however, by the frequency of each term in isolation: Borgatti's MDS solutions were centred on a core of prototypal, often-listed members of the domain in question, surrounded like the yolk of a fried egg by a halo of progressively less-salient members, frequency becoming a radial gradient while other axes were uninterpretable. At the other extreme one could assume that activation declines very rapidly with time, and include only direct adjacency in the calculation of similarity, accruing a point towards the association of two terms only if they appear consecutively in a list. However, this results in sparse data.

Henley's (1969) definition normalised ranking differences by the length of each list (Chan et al. 1993 and Aloia et al. 1996 also used normalising terms; also Prescott et al. 2006). Finally, in one extension of multidimensional scaling, the N points within a spatial model are adjusted to match 'three-way dissimilarities', which describe the relatedness of three items at a time, and are stored in a N-by-N-by-N matrix (e.g. Cox & Cox & Branca 1991). Definitions of list-based dissimilarity easily generalise to this approach.

Dissimilarities can also be obtained from text corpora by treating them as simple lists of words and calculating the lexical co-occurrence between pairs of words (Goñi et al. 2011; Lund & Burgess 1996).

CONCLUSIONS

Previous studies of colour linguistics used the listing or free-emission task as part of a standard field method for eliciting the basic colour terms in a given language (Davies & Corbett 1994a, 1994b; Sutrop 2002; Uusküla 2007; Vejdemo et al. 2015). In this context, the participants' lists are processed to yield a sali-

ence value for each term such as CSI or Smith's function. Listing data are also often collected in other domains of linguistic or anthropological interest such as body parts or kinship terms.

On its own, a discontinuity in the otherwise-exponential decline of CSI is not an infallible criterion of 'basicness'. An unequivocal BCT can slip down the rankings if its appearance in lists is reduced by the weakness of the cognitive associations connecting it to other BCTs, while conversely a hyponym or recessive near-synonym of a BCT can be elevated into the 'basic' side of the step if it is strongly linked to its dominant partner (this occurs for 'violet' in the English language results from Smith et al. 1995). In addition, the step is not universally obvious or abrupt (Davies & Corbett 1994b); it may continue over more than one interval and leave some terms partway down the steeper incline, requiring a subjective judgement whether they are BCTs or not. Cognitive salience is only one indicator, to be weighed with others.

Other research traditions have adopted the listing task for other objectives – notably anthropology and cognitive psychology (Borgatti 1999; Weller & Romney 1988), and clinical psychiatry where it serves a diagnostic purpose as the Noun Fluency Test. These other traditions inspired us to apply multidimensional scaling to listing data, resulting in spatial representations of colour terms in different languages.

A spatial map is a good fit to a list if the sum of successive distances within it – following a trajectory between points in the order those terms were selected – is relatively low, with sequential terms being spatially adjacent. It is worth noting that MDS solutions derived from overt judgements of similarity provided a good fit, in this sense, to *recalled* sequences of terms (Caramazza et al. 1976; Romney et al. 1993). The process of converting lists to adjacency estimates, and applying MDS to convert those in turn into language-specific maps, can be understood as a way of locating points to minimise this trajectory length, averaged across participants.

A recurring feature of the MDS solutions is the overall shape, reminiscent of a comet. Individual maps consist of a compact 'head' of first-used, highly-associated terms at one extreme of the 'priority' axis, while at the other axial extreme the lower-priority terms spread out across the second dimension. This shape is a natural outcome of listing behaviour. By definition, highly-salient terms all appear near the start of the lists, limiting the differences among their ranks. Conversely, less-salient terms can appear anywhere in a list, which allows larger ranking differences among them and requires more spread along the second axis. When Corbett and Davies (1997, Figure 9.2) used Correspondence Analysis to compare a range of different behavioural and text-based measures of basicness, they found that English colour terms were spread out across a similar geometrical 'map': salience was one axis, while a second axis expressed

secondary differences. In an attempt to eliminate 'salience' as a confounding factor in their calculation of list-based similarity, Aloia, Gourovitch, Weinberger and Goldberg (1996) and Chan, Butters, Paulsen, Salmon, Swenson and Maloney (1993) included terms to weight each subject's contribution. We feel it is easier to allow salience to emerge as its own separable dimension.

The second dimension captures a distinction between prototypal terms of 'real colour' and unchromatic, marginal or 'second-class' terms. Among primary basic terms, it separates 'black' and 'white' from the cardinal hues; 'grey' and often 'pink' and 'brown' from chromatic secondary terms; while among the non-basic terms it distinguishes the metallic sheens 'gold' and 'silver' from chromatic hyponyms like 'mauve'. A participant's chain of associations might jump between more- and less-basic chromatic terms, or between desaturated concepts.

For languages with sufficient data to include larger numbers of terms in the analysis – modified basics, or metonyms – the picture is complicated when these are inserted into two-dimensional maps (often as clumps of points, grouped together by common derivation from the same basic term). It may be that three-dimensional MDS solutions would reveal other cross-cultural commonalities.

Recent attention has turned beyond the inner circle of BCTs, to terms in common but not universal use (which might potentially become basic if the need for colour communication within a culture places enough emphasis on specificity) (Mylonas & MacDonald 2016; Lindsey & Brown 2014; Jraissati et al. 2012). *Turquoise* and German *türkis* have been mentioned as a possible incipient BCT (Zollinger 1984). However, a recent survey of American English (Lindsey & Brown 2014) found 'teal' to be the more common term for blue-green stimuli. The trend here, though with exceptions, was for 'turquoise' cognate terms to be linked with metallic sheens, suggesting that the concept is dominated by its non-chromatic aspects (perhaps emphasising the function of turquoise as a semi-precious component in jewellery).

In another example, Eessalu and Uusküla (2013) noted the relatively high salience of the equivalents of 'beige' in many of the present data sets; terms like Turkish *bej* and Hungarian *bézs* (also present in Hungarian as *drapp*). Again, these terms were coupled with achromatic terms such as 'grey', suggesting that the concept is not a chromatic one: 'beige' being an *absence* of decisive colour. In English, 'beige' does not feature among the 21 most salient terms in Smith, Furbee, Maynard, Quick, and Ross (1995). Sturges and Whitfield (1995) found 'beige' to be one of the three most frequent non-basic English terms for colour naming – the other two being 'cream' and 'turquoise' – and noted its value for identifying hues in a region of colour space that is poorly served by the 11 BCTs, and remote from their foci. The prominence of the term is a recent phenomenon: in a similar study eight years earlier (Boynton & Olson 1987), the use of 'beige' showed little consistency or popularity.

Distinctions among BCTs are also of interest. It seems natural to distinguish the four 'cardinal' terms plus white and black as 'primary' terms, and to regard the remaining BCTs as secondary: each derived from two primaries, combining them or occupying the borderland between them (Kay & McDaniel 1978). Some measures of basicness support this intuition, but not all (Corbett & Davies 1997; Sturges & Whitfield 1995). Many of the present CSI plots and MDS solutions display a gap between the primary and derived BCTs.

In several of the languages studied here, the sector of colour space covered by the English category 'blue' is split between a 'primary blue' term and a complementary 'secondary blue', both basic, though the level of lightness separating the two categories is not necessarily the same in each language. In Italian, for instance, *azzurro* and *celeste* 'light blue' are both candidates for basic status, although they are less inclusive than *blu* 'blue' or 'dark blue' (Paramei & Menegaz 2013; Bimler & Uusküla 2014). Lithuanian and Udmurt are two more examples, with 'light blue' terms *žydra* and *čagyr* respectively; contact with Russian may have encouraged their emergence (several strands of evidence support the basicness of *žydra*: Bimler & Uusküla, in preparation). In the present results, the secondary terms *azzurro*, *celeste*, Udmurt *čagyr* 'light blue' and Turkish *lacivert* 'dark blue' were all treated as chromatic concepts, while in contrast Lithuanian *žydra* 'light blue' had more in common with *pilka* 'grey'.

Similarities across language-specific MDS solutions point to a shared, cross-cultural pattern of associations among terms, structuring the sequence in which they are listed. The 'chunking' of terms indicates a shared system of conceptual attributes used to group them. In addition, dimensions of the MDS maps reflect conceptual themes at a higher level of abstraction. Note that this is not simply a consequence of the familiar cross-cultural regularities in the way that BCTs partition colour space into categories, and in the centroids and focal hues of these categories (i.e. the denotational use of the terms). This follows from the divergence between the maps and the spatial maps obtained when people rate the *perceptual* similarities among colour terms (from observation or from memory): listing associations access a more abstract aspect of 'similarity'. Perceptual opposites may be closely related at the conceptual level ('black' and 'white', 'purple' and 'orange').

ACKNOWLEDGEMENTS

Some of these data and versions of diagrams were previously reported to the Gruppo del Colore meeting in Genoa.

The whole data collection was founded by the Estonian Science Foundation grant no. 8168 (awarded to Mari Uusküla) and Estonian Ministry of Education and Research project no. SF0050037s10 (awarded to Urmas Sutrop). The authors are thankful to all data collectors.

REFERENCES

Aloia, Mark S. & Gourovitch, Monica L. & Weinberger, Daniel R. & Goldberg, Terry E. 1996. An Investigation of Semantic Space in Patients with Schizophrenia. *Journal of International Neuropsychological Society*, Vol. 2, No. 4, pp. 267–273. http://dx.doi.org/10.1017/S1355617700001272.

Berlin, Brent & Kay, Paul 1991 [1969]. *Basic Color Terms: Their Universality and Evolution*. Berkeley & Los Angeles & Oxford: University of California Press.

Bimler, David & Uusküla, Mari 2014. "Clothed in Triple Blues": Sorting Out the Italian Blues. *Journal of the Optical Society of America*, Vol. 31, No. 4, pp. A332–A340. http://dx.doi.org/10.1364/JOSAA.31.00A332.

Borgatti, Stephen P. 1999. Elicitation Techniques for Cultural Domain Analysis. In: J. Schensul & M. D. LeCompte (eds.) *Ethnographer's Toolkit*, Vol. 3. Walnut Creek: AltaMira Press, pp. 115–151.

Boynton, Robert M. & Conrad X. Olson 1987. Locating Basic Colors in the OSA Space. *Color Research & Application*, Vol. 12, No. 2, pp. 94–105. http://dx.doi.org/10.1002/col.5080120209.

Caramazza, Alfonso & Hersh, Harry & Torgerson, Warren S. 1976. Subjective Structures and Operations in Semantic Memory. *Journal of Verbal Learning and Verbal Behavior*, Vol. 15, No. 1, pp. 103–117. http://dx.doi.org/10.1016/S0022-5371(76)90011-6.

Chan, Agnes S. & Butters, Nelson & Paulsen, Jane S. & Salmon, David P. & Swenson, Michael R. & Maloney, Laurence T. 1993. An Assessment of the Semantic Network in Patients with Alzheimer's Disease. *Journal of Cognitive Neuroscience*, Vol. 5, No. 2, pp. 254–261. doi:10.1162/jocn.1993.5.2.254.

Collins, Allan M. & Loftus, Elizabeth F. 1975. A Spreading-Activation Theory of Semantic Processing. *Psychological Review*, Vol. 82, No. 6, pp. 407–428. http://dx.doi.org/10.1037/0033-295X.82.6.407.

Corbett, Greville C. & Davies, Ian R. L. 1997. Establishing Basic Color Terms: Measures and Techniques. In: C. L. Hardin & L. Maffi (eds.) *Color Categories in Thought and Language.* Cambridge: Cambridge University Press, pp. 197–223.

Cox, Trevor F. & Cox, Michael A. A. & Branco, Joao A. 1991. Multidimensional Scaling for *n*-Tuples. *British Journal of Mathematical & Statistical Psychology*, Vol. 44, No. 1, pp. 195–206. http://dx.doi.org/10.1111/j.2044-8317.1991.tb00955.x.

Davies, Ian R. L. & Corbett, Greville C. 1994a. A Statistical Approach to Determining Basic Color Terms: An Account of Xhosa. *Journal of Linguistic Anthropology*, Vol. 4, No. 2, pp. 175–193. http://dx.doi.org/10.1525/jlin.1994.4.2.175.

Davies, Ian R. L. & Corbett, Greville C. 1994b. The Basic Color Terms of Russian. *Linguistics*, Vol. 32, No. 1, pp. 65–89. http://dx.doi.org/10.1515/ling.1994.32.1.65.

Davies, Ian R. L. & Corbett, Greville C. & Margalef, José Bayo 1995. Colour Terms in Catalan: An Investigation of Eighty Informants, Concentrating on the Purple and Blue Regions. *Transactions of the Philological Society*, Vol. 93, No. 1, pp. 17–49. http://dx.doi.org/10.1111/j.1467-968X.1995.tb00435.x.

Eessalu, Martin & Uusküla, Mari 2013.The Special Case of *Beige*: A Cross-Linguistic Study. In: M. Rossi (ed.) *Colour and Colorimetry: Multidisciplinary Contributions,* Vol. IXB. Maggioli: Rimini, pp. 168–176. Available at https://www.researchgate.net/publication/264461470_The_special_case_of_beige_A_cross-linguistic_study, last accessed on May 20, 2016.

Fletcher, Robert 1980. *The City University Colour Vision Test.* 2nd ed. London: Keeler.

Friendly, Michael L. 1977. In Search of the M-Gram: The Structure of Organization in Free Recall. *Cognitive Psychology*, Vol. 9, No. 2, pp. 188–249. http://dx.doi.org/10.1016/0010-0285(77)90008-1.

Friendly, Michael L. 1979. Methods for Finding Graphic Representations of Associative Memory Structures. In: C. R. Puff (ed.) *Memory Organization and Structure*. New York: Academic Press, pp. 85–129.

Goñi, J. & Arrondo, G. & Sepulcre, J. & Martincorena, I. & Vélez de Mendizabal, N. & Corominas-Murtra, B. & Bejarano, B. & Ardanza-Trevijano, S. & Peraita, H. & Wall, D. P. & Villoslada, P. 2011. The Semantic Organization of the Animal Category: Evidence from Semantic Verbal Fluency and Network Theory. *Cognitive Processing*, Vol. 12, No. 2, pp. 183–196. DOI: 10.1007/s10339-010-0372-x.

Harkness, Sarah 1973. Universal Aspects of Learning Color Codes: A Study in Two Cultures. *Ethos*, Vol. 1, No. 2, pp. 175–200. DOI: 10.1525/eth.1973.1.2.02a00030.

Henley, Nancy M. 1969. A Psychological Study of the Semantics of Animal Terms. *Journal of Verbal Learning and Verbal Behavior*, Vol. 8, No. 2, pp. 176–184. http://dx.doi.org/10.1016/S0022-5371(69)80058-7.

Hippisley, Andrew R. 2001. Basic BLUE in East Slavonic. *Linguistics,* Vol. 39, No. 1, pp. 151–179. http://dx.doi.org/10.1515/ling.2001.003.

Ishihara, Shinobu 2008. *Ishihara's Tests for Colour Deficiency*. 24 Plates Edition. Tokyo: Kanehara Trading Inc.

Jraissati, Yasmina & Wakui, Elley & Decock, Lieven & Douven, Igor 2012. Constraints on Colour Category Formation. *International Studies in the Philosophy of Science*, Vol. 26, No. 2, pp. 171–196. http://dx.doi.org/10.1080/02698595.2012.703479.

Kay, Paul & McDaniel, Chad K. 1978. The Linguistic Significance of the Meanings of Basic Color Terms. *Language*, Vol. 54, No. 3, pp. 610–646. http://dx.doi.org/10.1353/lan.1978.0035.

Kerttula, Seija 2007. Relative Basicness of Color Terms: Modeling and Measurement. In: Robert E. MacLaury & Galina V. Paramei & Don Dedrick (eds.) *Anthropology of Color.* Amsterdam & Philadelphia: John Benjamins Publishing Company, pp. 151–169. DOI: 10.1075/z.137.11ker.

Lindsey, Delwin T. & Brown, Angela M. 2014. The Color Lexicon of American English. *Journal of Vision*, Vol. 14, No. 2, pp. 1–25. http://dx.doi.org/10.1167/14.2.17.

Lund, Kevin & Burgess, Curt 1996. Producing High-Dimensional Semantic Spaces from Lexical Co-Occurrence. *Behavior Research Methods, Instruments, & Computers*, Vol. 28, No. 2, pp. 203–208. http://dx.doi.org/10.3758/BF03204766.

Mylonas, Dimitris & MacDonald, Lindsay 2016. Augmenting Basic Colour Terms in English. *Color Research & Application*, Vol. 41, No. 1, pp. 32–42. http://dx.doi.org/10.1002/col.21944.

Mylonas, Dimitris & Paramei, Galina V. & MacDonald, Lindsay 2014. Gender Differences in Colour Naming. In: Wendy Anderson & Carole P. Biggam & Carole Hough & Christian Kay (eds.) *Colour Studies: A Broad Spectrum*. Amsterdam & Philadelphia: John Benjamins Publishing Company, pp. 225–239. DOI: 10.1075/z.191.15myl.

Özgen, Emre & Davies, Ian R. L. 1998. Turkish Color Terms: Tests of Berlin and Kay's Theory of Color Universals and Linguistic Relativity. *Linguistics,* Vol. 36, No. 5, pp. 919–956. http://dx.doi.org/10.1515/ling.1998.36.5.919.

Paramei, Galina V. 2005. Singing the Russian Blues: An Argument for Culturally Basic Color Terms. *Cross-Cultural Research*, Vol. 39, No. 1, pp. 10–38. http://dx.doi.org/10.1177/1069397104267888.

Paramei, Galina V. 2007. Russian 'Blues': Controversies of Basicness. In: Robert E. MacLaury & Galina V. Paramei & Don Dedrick (eds.) *Anthropology of Color.* Amsterdam & Philadelphia: John Benjamins Publishing Company, pp. 75–106. DOI: 10.1075/z.137.07par.

Paramei, Galina V. & Menegaz, Gloria 2013.'Italian Blues': A Challenge to the Universal Inventory of Basic Colour Terms. In: Maurizio Rossi (ed.) *Colour and Colorimetry: Multidisciplinary Contributions, Vol. IX B*, pp. 164–167. Rimini: Maggioli S.p.A.

Paulsen, Jane S. & Romero, Ramon & Chan, Agnes & Davis, Amy V. & Heaton, Robert K. & Jeste, Dilip V. 1996. Impairment of the Semantic Network in Schizophrenia. *Psychiatry Research*, Vol. 63, No. 2–3, pp. 109–121. http://dx.doi.org/10.1016/0165-1781(96)02901-0.

Pranaitytė, Simona 2011. Põhivärvinimed leedu keeles. [The Basic Colour Terms of Lithuanian.] In: Mari Uusküla & Urmas Sutrop (eds.) *Värvinimede raamat.* [A Book of Colour Terms.] Tallinn: Eesti Keele Sihtasutus, pp. 271–300. Available at http://www.academia.edu/23710383/V%C3%A4rvinimede_raamat, last accessed on May 23, 2016.

Prescott, Tony J. & Newton, Lisa D. & Mir, Nusrat U. & Woodruff, Peter W. R. & Parks, Randolf W. 2006. A New Dissimilarity Measure for Finding Semantic Structure in Category Fluency Data with Implications for Understanding Memory Organization in Schizophrenia. *Neuropsychology*, Vol. 20, No. 6, pp. 685–699. http://dx.doi.org/10.1037/0894-4105.20.6.685.

Rätsep, Kaidi 2011. Preliminary Research on Turkish Basic Colour Terms with an Emphasis on Blue. In: C. P. Biggam & C. A. Hough & C. J. Kay & D. R. Simmons (eds.) *New Directions in Colour Studies.* Amsterdam & Philadelphia: John Benjamins Publishing Company, pp. 133–145.

Rätsep, Kaidi 2013. Some Remarks on Gender Differences in Turkish Colour Vocabulary. In: Balázs Surányi (ed.) *Proceedings of the Second Central European Conference in Linguistics for Postgraduate Students.* Second Central European Conference in Linguistics for Postgraduate Students (CECIL'S 2), Pázmány Péter Catholic University, Piliscsaba, Hungary, on 24–25 August 2012. Budapest: Pázmány Péter Catholic University, pp. 219–229.

Regier, Terry & Kay, Paul & Cook, Richard S. 2005. Universal Foci and Varying Boundaries in Linguistic Color Categories. In: Bruno G. Bara & Lawrence Barsalou & Monica Bucciarelli (eds.) *Proceedings of the 27th Annual Conference of the Cognitive Science Society (CogSci 2005), July 21–23, Stresa, Italy*, pp. 1827–1832. Available at http://csjarchive.cogsci.rpi.edu/proceedings/2005/docs/p1827.pdf, last accessed on May 23, 2016.

Roberson, Debi & Davidoff, Jules & Davies, Ian R. L. & Shapiro, Laura R. 2004. The Development of Color Categories in Two Languages: A Longitudinal Study. *Journal of Experimental Psychology: General*, Vol. 133, No. 4, pp. 554–571. http://dx.doi.org/10.1037/0096-3445.133.4.554.

Romney, A. Kimball & Brewer, Devon D. & Batchelder, William H. 1993. Predicting Clustering from Semantic Structure. *Psychological Science*, Vol. 4, No. 1, pp. 28–34. http://dx.doi.org/10.1111/j.1467-9280.1993.tb00552.x.

Ryabina, Elena 2009. Sex-Related Differences in the Colour Vocabulary of Udmurts. *WEBFU: Wiener elektronische Beiträge des Instituts für Finno-Ugristik*. Available at http://www.univie.ac.at/webfu/texte/12Ryabina.pdf, last accessed on May 23, 2016.

Ryabina, Elena 2011. Differences in the Distribution of Colour Terms in Colour Space in the Russian, Udmurt and Komi Languages. *ESUKA-JEFUL*, Vol. 2, No. 2, pp. 191–213. Available at http://jeful.ut.ee/public/files/Ryabina%20191-213.pdf, last accessed on May 23, 2016.

Shepard, Roger N. 1974. Representation of Structure in Similarity Data: Problems and Prospects. *Psychometrika*, Vol. 39, No. 4, pp. 373–421. http://dx.doi.org/10.1007/BF02291665.

Smith, J. Jerome 1993. Using ANTHROPAC 3.5 and a Spreadsheet to Compute a Free-List Salience Index. *Field Methods*, Vol. 5, No. 3, pp. 1–3. doi:10.1177/1525822X9300500301.

Smith, J. Jerome & Furbee, Louanna & Maynard, Kelly & Quick, Sarah & Ross, Larry 1995. Salience Counts: A Domain Analysis of English Color Terms. *Journal of Linguistic Anthropology*, Vol. 5, No. 2, pp. 203–216. http://dx.doi.org/10.1525/jlin.1995.5.2.203.

Storm, Christine 1980. The Semantic Structure of Animal Terms: A Developmental Study. *International Journal of Behavioral Development*, Vol. 3, No. 4, pp. 381–407. http://dx.doi.org/10.1177/016502548000300403.

Sturges, Julia & Whitfield, T. W. Allan 1995. Locating Basic Colours in the Munsell Space. *Color Research and Application*, Vol. 20, No. 6, pp. 364–376. http://dx.doi.org/10.1002/col.5080200605.

Sutrop, Urmas 2001. List Task and a Cognitive Salience Index. *Field Methods*, Vol. 13, No. 3, pp. 263–276. http://dx.doi.org/10.1177/1525822X0101300303.

Sutrop, Urmas 2002. *The Vocabulary of Sense Perception in Estonian: Structure and History*. Frankfurt am Main & Berlin & Bern & Bruxelles & New York & Oxford: Peter Lang.

Uusküla, Mari 2007. The Basic Colour Terms of Finnish. *SKY Journal of Linguistics* 20, pp. 367–397. Available at http://www.linguistics.fi/julkaisut/sky2007.shtml, last accessed on May 23, 2016.

Vejdemo, Susanne & Levisen, Carsten & van Scherpenberg, Cornelia & Beck, Þórhalla Guðmundsdóttir & Næss, Åshild & Zimmermann, Martina & Stockall, Linnaea & Whelpton, Matthew 2015. Two Kinds of Pink: Development and Difference in Germanic Colour Semantics. *Language Sciences*, Vol. 49, pp. 19–34. http://dx.doi.org/10.1016/j.langsci.2014.07.007.

Weller, Susan C. & Romney, A. Kimball 1988. *Systematic Data Collection*. Newbury Park & London & New Delhi: SAGE publications.

Zollinger, Heinrich 1984. Why Just Turquoise? Remarks on the Evolution of Color Terms. *Psychological Research*, Vol. 46, No. 4, pp. 403–409. http://dx.doi.org/10.1007/BF00309072.

http://dx.doi.org/10.7592/FEJF2016.64.sandford

COLOR ENTRENCHMENT IN MIDDLE-SCHOOL ENGLISH SPEAKERS: COGNITIVE SALIENCE INDEX APPLIED TO COLOR LISTING

Jodi L. Sandford

Abstract: This study presents results of two questionnaires posed to English speaking middle-school students to verify the level of color term entrenchment and color prototypes at the age of 12. The methods included color listing and informant introspection on a color prototype linguistic construction. Listing techniques have long been used to identify basic categories and prototypically relevant linguistic items. In this case Sutrop's Cognitive Salience Index (2001) served to reveal the facilitation of retrieval of concepts in long-term memory thus allowing us to evince the degree of entrenchment and salience of the given color term. Participant introspection in conjunction with conceptual salience analysis (Talmy 2000, 2005) regarding prototypical items (Rosch 1975, 1978, 1983) was also employed to identify what items are actually associated prototypically with colors at this age. The results are compared to adult color listings and prototypes (for both English and Italian speakers). Divergence is significant both in regard to the Cognitive Salience Index and within group judgment of the color prototypes. Further details about the subordinate color term choices and the agreement on prototypes reveal the conventionalized linguistic color associations made by this specific group of north-west American middle-school students.

Keywords: basic color terms, cognitive salience index, color listing, color prototypes, entrenchment, salience

INTRODUCTION

This paper illustrates the results of two questionnaires presented to middle-school children regarding color conceptualization in English. One questionnaire involved color term listing; a method employed to reveal general color term entrenchment. The other questionnaire asked for informant introspection on color prototypes in relation to the six primary basic color terms (BCT's) (Berlin & Kay 1991; Kay et al. 2009). Color term listing according to the cognitive semantic approach reveals the degree of entrenchment of specific colors. This approach is concerned with investigating the relationship between experience, the conceptual system, and the semantic structure encoded by language. More

precisely it is involved in understanding the structure of language, as it emerges from knowledge representation, and the conceptualization in language that is reflected in meaning construction (Evans 2007: 26). As Schmid explains, if words of a language represent the conceptualizations that have been fossilized by convention in a speech community, then their "entrenchment" is the strength or "degree to which the formation and activation of a cognitive unit is routinized and automated" (2007: 118). Croft and Cruse (2004) identify "entrenchment" with the attainment of some sort of default status. Moreover, Langacker (2008: 21) specifies that entrenchment, or unit status, pertains to individual speakers' usage, whereas conventionality pertains to a community of speakers. That is to say that the more we use a specific word, the more automatically we will continue to use it and the more it will become cognitively entrenched for both the speaker and the community of speakers.

By using the Cognitive Salience Index (CSI) developed by Sutrop (2001) I was able to identify the most salient color words in the color lists and rank the degree of entrenchment. The rationale behind the CSI is that there is a correlation between entrenchment of terms and their accessibility for a higher number of the participants. The CSI is the result of the equation CSI = F/ (N X MP), where F (relative frequency of a term) is divided by (N (number of informants) multiplied by MP (the mean position of the term)). In so doing, the terms that tend to be retrieved more quickly (and therefore tend to be listed among the first entries) have a higher MP, which will result in a higher CSI value. The CSI yields a value between 0 and 1, with greater values corresponding to greater cognitive salience. The cognitive salience index yields comparable results across studies since it does not depend on the length of the individual lists (Sutrop 2001: 267). Numerous researchers have used the Cognitive Salience Index as a successful means of establishing a reliable ranking of items (see, for example, Kuehnast et al. 2014; Sandford 2015; Sutrop 2000, 2001, 2002: Uusküla & Sutrop 2007; Uusküla 2007, 2008).

A color term is salient if it is readily elicitable, occurs in the idiolects of most speakers, and is used consistently by individuals and with a high degree of consensus among individuals (Hardin & Maffi 1997: 4). Schmid defines the notion of salience that "may thus denote both a temporary activation state of mental concepts (cognitive salience) and an inherent and consequently more or less permanent property of entities in the real world (ontological salience). […] As a result, cognitive events related to the processing of ontologically salient entities will occur more frequently and lead to earlier entrenchment of corresponding cognitive units, or concepts. […] On the other hand deeply entrenched cognitive units are more likely to become cognitively salient than less well entrenched ones" (2007: 120).

This is how color listing and color term recall may function as an indication of salience, both cognitive and ontological, in that the informants will reveal exactly those words that are used more frequently by themselves and by the linguistic community they are part of. And at the same time these lists reveal, according to their young age, the degree of entrenchment and salience that color words have for this group of young people in a multi ethnical American community.

Listing techniques have long been used to identify basic categories and prototypically relevant linguistic items. See for example the work by Corbett and Davies (1997: 203–205; 1995) and Davies and Corbett (1995, 1994) who used color name listing to evaluate the correlation between frequency and top of the list ranking as the measure of basicness. Listing is used to reveal exactly the facilitation of retrieval of concepts in long-term memory thus allowing us to evince the degree of entrenchment and salience of the given category items.

In the second questionnaire I asked the informants to write down prototypical items, in keeping with Talmy's (2000) indications on introspection as a methodological research tool. I approached this by giving them the expression, or construction[1], "as [color] as X". Here [color] stands for one of the 6 primary BCT's and X represents the blank that the informants were asked to fill with the first word that came to mind in relation to that color. As Geeraerts claims, "when informants are asked to enumerate the members of a category, typical members are more often named than marginal ones" (2009: 187). These typical members are known as prototypes, or cognitive representations of best exemplars. Categories are based on such shared mental concepts. Moreover, by prototypes I mean "the clearest cases of category membership defined operationally by people's judgments of goodness of membership in the category", in keeping with Rosch (1978: 36).

Rosch subsumed the implications of this type of research in 1973. She affirms that the nature of cognitive representations of semantic categories has direct relevance to two important areas of inquiry. One concerns the structure of categories and concepts and has implications for the way in which concepts may be researched. The other area delves into the nature of mental representations, as proposed in the cognitive commitment of cognitive linguistics (see Lakoff 1990 [1987]). These two areas of investigation, the structure of categories and concepts, and their mental representation, are pertinent to this analysis. Furthermore, the commitment, taken into consideration in this approach, represents a dedication to characterizing general principles that apply to all aspects of human language and how they emerge from a common set of human cognitive abilities, and to providing a characterization of the general principles for language that harmonize with what is known about the mind and brain

from other disciplines. Rosch emphasizes how our ability to process the frame of color is emblematic of this type of revelation of general principles. A color term reveals not only a linguistic prototypical reference point that corresponds to a visual focal color, but also the associated object or prototypical associate itself.

> *There is now considerable evidence that color categories are processed by the human mind (learned, remembered, denoted, and evolved in languages) in terms of their internal structure; color categories appear to be represented in cognition not as a set of criterial features with clear-cut boundaries but rather in terms of a prototype (the clearest cases, best examples) of the category, surrounded by other colors of decreasing similarity to the prototype and of decreasing degree of membership.* (Rosch 1975: 193)

In this study we are referring both to the most entrenched color word/concepts for these 12 year olds, and to the central best example of the focal color itself in terms of the object that gives us the cognitive reference to that visual focal color.

Current research, since Berlin and Kay (1991 [1969]) and Kay et al. (2009), generally considers there to be eleven BCTs in English, that fit into the evolutionary hierarchy of BLACK and WHITE, RED, YELLOW and GREEN, BLUE, GREY and BROWN, PURPLE, PINK, ORANGE (see Biggam 1997; Casson 1997; Dedrick et al. 2005; Sandford 2012). These colors plus one were used as the basis for this analysis.

METHOD

Informants

The informants included 29 students, 16 females and 13 males from the Explorer Middle School in Everett, Washington. The class of students was a special group of highly gifted or "honor" students. The average age was 11.6 years old, the oldest 12 and the youngest 11. All of the students were native American English speakers. Observing the ethnic mixture of the students, I asked them if they spoke another language at home and 24 students raised their hands. So it is relevant to note that a majority of these advanced students from this suburban neighbourhood had some sort of bilingual input. Their parents were from over 7 different countries: USA, China, Korea, India, Italy, Central America, and South America.

Procedure

The students received a short briefing explaining that we were going to be talking about color, but before discussing the matter, so as not to prime their responses, I wanted them to list color words by writing them down on the piece of paper I had put on their desks. I asked them to fold the paper into three parts horizontally, and write as many color words as they could in each column. I explained that they would have a minute for each column for a total of three minutes and that I would tell them when to move to each column.

After this the students were asked to fill in a separate paper questionnaire. Each questionnaire had 12 phrases listed on the paper, two for each of the six primary BCTs: Black, White, Red, Yellow, Green, Blue. There was a blank next to each phrase "as [color] as X", for example, as black as …… , and the informants were instructed to write the first word that they associated with the color in the blank.

These data were also analyzed using the Cognitive Salience Index (Sutrop 2001) taking into account two important aspects: term frequency and mean position (see above). The cognitive salience index was calculated to find the CSI rank of each color word in the lists of colors provided by the informants.

RESULTS

The students put together a list of 163 different color names. The total list included 730 items. The longest list included 39 colors and the shortest list 15 colors (SD 7.34). An average of 25 colors total per informant, 14 colors average the first minute, 6 colors average the second minute, and 5 colors average the third minute. Of the 163 different colors, only 16 colors were named by half of the informants, another 17 colors were named by 13 to 6 informants. This means that 130 of the colors were named by fewer than 5 informants (4 colors were named by 5 informants, 10 colors by 4, 11 colors by 3, 18 colors by 2, and 87 colors by 1). Appendix 1 shows the 25 colors with the highest CSI. All of the colors are listed in note 2[2].

Listing task and CSI

Table 1 shows the top 12 CSI colors: Red, Orange, Yellow, Blue, Green, Purple, Magenta, Black, Pink, White, Brown, and Grey. The CSI is considered the measure of the most prominent/salient colors named by a group of people.

Table 1. *The Twelve Top CSI Colors.*

COLOR	MP	MP RANK	FREQ	% FREQ	FREQ RANK	CSI (S = F / (N X MP)	CSI RANK
Red	4	1	28	97	3	0.2414	**1**
Orange	5	2	29	100	1	0.2000	**2**
Yellow	5	3	28	97	4	0.1931	**3**
Blue	5	4	27	93	5	0.1862	**4**
Green	5	5	27	93	6	0.1862	**5**
Purple	8	8	25	86	9	0.1078	**6**
Magenta	9	9	27	93	7	0.1034	**7**
Black	11	19	26	90	8	0.0815	**8**
Pink	13	29	29	100	2	0.0769	**9**
White	12	24	25	86	10	0.0718	**10**
Brown	14	34	25	86	11	0.0616	**11**
Grey	14	35	24	83	12	0.0591	**12**

I calculated the Index for all of the colors and ranked them accordingly. The colors that had the highest *Frequency* ranks were *Orange*, *Pink*, *Red*, *Yellow*, *Blue*, *Green*, *Magenta*, *Black*, *Purple*, *White*, *Brown*, and *Grey*. The colors *Black, White, Brown*, and *Grey* do not change rank between the two measures: frequency and CSI. *Orange, Red*, and *Yellow* rank in the top four in both rankings, but the Mean Position ranking changes. *Yellow*, *Blue*, and *Green* stay in the same order for both rankings.

The main shifts in rank are *Pink* and *Purple*, and *Magenta* to a lesser degree. That is, *Pink* was listed by all of the informants but only later in the lists, *Purple* was listed by fewer informants but much earlier in the lists. *Magenta* was the only non-BCT that was listed by almost all the informants fairly early in the lists. This is why Sutrop's Index is so informative; by including all the different aspects of information in the calculation we are able to give a more accurate idea of the individual entrenchment and the group response.

Prototype listing task

The two most common objects, or prototypes, that emerged for each color construction list are: Night and Pitch for *Black*; Paper and Snow for *White*; Blood and Rose for *Red*; Sun and Dandelions for *Yellow*; Grass and Leaves for *Green*; Sky and Ocean for *Blue*. Table 2 lists the 4 most frequent associates. In this case I applied the CSI, though the variation in MP is limited since there were only two items per informant. It still became pertinent to see if the prototypes emerged first or second. For example, in the case of *Green* Emeralds were listed only twice but in both occasions first, and Frogs were listed 3 times but always second, which resulted in a lower CSI.

Table 2. *Four Top CSI Color Prototypes per Color for 29 Informants.*

COLOR	**Prot. 1**	**#**	**CSI**	**Prot. 2**	**#**	**CSI**	**Prot. 3**	**#**	**CSI**	**Prot. 4**	**#**	**CSI**
BLACK	Night	17	0.5205	Pitch	6	0.1379	Space	5	0.1232	Coal	4	0.0788
WHITE	Paper	19	0.4614	Snow	9	0.2333	Cloud(s)	7	0.1412	Light	3	0.0788
RED	Blood	17	0.5010	Rose(s)	8	0.1692	Fire	6	0.1379	Apple	5	0.0958
YELLOW	Sun	18	0.5592	Dande-lions	4	0.1103	Lemons	4	0.0920	Ba-nana	4	0.0788
GREEN	Grass	20	0.6270	Leaves	9	0.1744	Emeralds	2	0.0690	Frogs	3	0.0619
BLUE	Sky	19	0.5000	Ocean	14	0.3376	Sea	5	0.1232	Water	4	0.0690

DISCUSSION

It is pertinent and in keeping with basic color theory that the 11 basic color terms in English [*Black*, *White*, *Red*, *Green*, *Yellow*, *Blue*, *Brown*, *Grey*, *Purple*, *Orange*, *Pink*] are most prominent for the 12 year olds. *Magenta*, rank 7th in CSI, is the odd one in the list. It is clearly prominent for the middle-school students. An explanation may be tied to the use of computers and the division between the RGB (red, green, blue) autoluminant system that is contrasted to the CYM (cyan, yellow, magenta) pigment system. *Cyan* was also listed, but only rank 36th in CSI. It may, on the other hand, have to do with their studies. The class was studying the middle ages and there may have been some mention of the pigments that were commonly used at that time. This came to mind since they also listed *Indigo*, rank 13th CSI. In this case it would suggest that *Magenta* may be shifting towards a BCT, taking the position that was previously held by

Pink. As Desgrippes states, "the cognitive representation of a color is dynamic: it can evolve with diachronic language variation or with language shift, and both older and newer representations remain retrievable depending on the task at hand" (2013). This may be the case for the middle-school students who have been exposed to advanced color terminology and in this circumstance, being highly habituated to recall of vocabulary, able to access numerous sophisticated terms especially those that fit the scholastic context.

It is further relevant to see that, after the 12 most prominent colors (Table 1), of the following 12 colors of the list of the 25 highest CSI measure (Appendix 1) 6 colors are types of blue, 3 are types of purple, the other 3 are non-colors, or low in saturation, *Gold*, *Silver*, and *Tan*. Moreover, the high CSI of the BCTs positioned the warm colors *Red*, *Orange*, and *Yellow* first over the cool colors *Blue*, *Green*, and *Purple*. These aspects would appear to be indicative of the initial perceptive salience of warm colors for the 12 year olds on one hand, and on the other to reveal the salience of a significant variation in blue color terms.

Of the total list of 163 colors it is possible to group them according to the form and content. Each group is presented in alphabetical order. Firstly, there are 6 questionable color names: *Clear*, *Neon indigo*, *Neon black*, *Neon brown*, *Rainbow*, and *UV colors.* Of the subordinate (non-basic) color names, there are 9 pigment names: *Chinese white*, *Indigo*, *Lamp black*, *Prussian Blue*, *Rose Madder*, *Russet*, *Siena*, *Ultramarine*, *Umber*; 8 food names: *Burgundy*, *Caramel*, *Chartreuse*, *Cream*, *Raspberry*, *Peach*, *Salmon*, *Strawberry*; 7 materials used for jewellery: *Amber*, *Aquamarine*, *Coral*, *Ebony*, *Emerald*, *Jade*, *Turquoise*; 5 metal names: *Bronze*, *Iron*, *Gold*, *Rust*, *Silver*; 5 flowers: *Fuchsia*, *Lavender*, *Lilac*, *Periwinkle*, *Rose*; 2 trees/wood: *Hazel*, *Mahogany*; 2 materials: *Brick*, *Earth*; *and* 4 other: *Egg shell*, *Jeans*, *Smokey*, *Sunset*.

Then there is a group of compound color terms, of which only 1 "Bright" color, 1 "Deep" color, 4 "Pastel" colors, 7 "Neon" colors, 9 "Dark" colors, and 11 "Light" colors. They are: 17 Blues (*Baby blue*, *Dark blue*, *Green-blue*, *Electric blue*, *Light blue*, *Midnight blue*, *Navy blue*, *Neon blue*, *Non-photo blue*, *Ocean blue*, *Pastel blue*, *Powder blue*, *Robin egg blue*, *Royal blue*, *Sea blue*, *Sky blue*, *Vibrant blue*); 16 Greens (*Dark green*, *Blue green*, *Faded green*, *Foam-green*, *Forest green*, *Grass green*, *Hunter green*, *Light green*, *Lime green*, *Mint green*, *Neon green, Pastel green*, *Pea green*, *Yellowish green*, *Yellow green*, *Sea green*); 13 Yellows (*Bright yellow*, *Bumblebee yellow*, *Butter yellow*, *Dark yellow*, *Fluorescent yellow*, *Indian yellow*, *Lemon yellow*, *Light yellow*, *Mustard yellow*, *Neon yellow*, *Orange yellow*, *Pastel yellow*, *Sunshine yellow*, and 2 variations *Rose-gold* and *Golden*); 10 Pinks (*Carnation pink*, *Coral pink*, *Electric pink*, *Hot-pink*, *Light pink*, *Neon pink*, *Pastel pink*, *Salmon pink*, *Soft pink*, *Tickle-me-pink*); 7 Reds (*Blood red*, *Dark red*, *Light red*, *Neon red*, *Orange-red*, *Peach red*, *Violet red*);

7 Purples (*Dark purple*, *Dark violet*, *Light purple*, *Light violet*, *Neon purple*, *Red violet*, *Royal violet*); 7 Oranges (*Yellow orange*, *Dark orange*, *Deep orange*, *Light orange*, *Neon orange*, *Red orange*, *Sunset orange*); 6 Greys (*Blue grey*, *Cool grey*, *Dark grey*, *French grey*, *Gravel Grey*, *Light grey*); 3 Browns (*Chocolate brown*, *Light brown*, *Reddish-brown*); 2 Indigos (*Dark indigo*, *Light indigo*); and 1 White (*Off-white*). The remaining 13 subordinate monolexemic color terms included: *Aqua*, *Azure*, *Beige*, *Cerulean, Crimson*, *Cyan*, *Magenta*, *Maroon*, *Navy*, *Scarlet*, *Tan*, *Teal*, *Violet* (6 blues, 4 reds, 2 browns, and 1 purple). This leaves the remaining 11 BCTs: *Black*, *Blue*, *Green*, *Grey*, *Brown*, *Orange*, *Pink*, *Purple*, *Red*, *Yellow*, *White*.

The students demonstrated a large vocabulary that reflected crayon and paint names. Some of the students were also aware of how to compound color words by using adjectives, like *neon-yellow*, *pastel-yellow*, *light-yellow*, *dark-yellow*, or *bright yellow*, or the colors themselves, for example, *orange yellow* or *yellow orange*. So the students used both long-term memory of stable references, objects with color names written on them, and linguistic mechanisms to be able to list as many names as they could.

Generally speaking, the age range considered for the onset of color term establishment reliability acquired of the first focal colors by 3 years and brown and grey 6–9 months later (Pitchford & Mullen 2002), around 4 years, Moreover, Pitchford and Mullen revealed that "of the 11 basic colors, brown and grey were the two least preferred colors, suggesting a developmental link between color preference and color term acquisition" (Pitchford 2006: 330). It is relevant to note that by 12 these children responses not only rank Brown and Grey lowest, reflecting the same type of preference; they also have a fully developed color term vocabulary.

When I compared these CSI ranking results to color name listings put together from a small group of middle aged American adults who attended a color workshop and a large group of Italian University students (for corresponding CSI see Appendix 1); fundamentally the eleven basic color terms emerged constantly at the top of the rank. The order varies and implies a different level of cognitive salience especially between the English children and the adults. Table 3 shows the first 12 plus 3 rankings with the SD of the ranking that resulted according to the CSI. In order to make the lists comparative I used a *Fuchsia* for English adults, and *Fucsia* for Italian adults as a close perceptive equivalent of *Magenta* that the English middle-school students listed 7th CSI rank. I deemed this to be legitimate since Fuchsia is defined as "magenta in color" in http://www.word reference.com/enit/fuchsia. However, the middle-school kids also named *Fuchsia* separately, CSI rank 23 on the list, though the adults did not mention *Magenta* (see Appendix 1). *Tan* for English students

and adults and was compared with *Beige* for Italian adults. The 3 rankings in Table 3 are listed according to the SD, from the lowest to the highest. This is to better illustrate the similarities that emerge.

The two English groups have the same CSI rank only for *Purple*. The two Adult groups have the same CSI rank for *Pink* and *Green*. The Middle school students and the Italian university students have the same CSI rank for *Red*, *Blue*, *Grey*, and *Brown*. It is hard to say how the specific color position in the list tells us something more specific. The order of approximate cognitive entrenchment results ranking across the three groups again according to SD as: *Red*, *Pink*, *Green*, *Purple*, *Blue*, *Black*, *Orange*, *White*, *Grey*, *Yellow*, and *Brown*. The ranking would seem to have positivity bias to it, where the top colors are associated more with positive metonymies and metaphors the higher on the list (see Sandford 2012, forthcoming).

Table 3. *General color ranking by CSI results for top 12 color listing (plus 3) and SD rank.*

Color	English Middle school	English Adults	Italian Adults	SD
red	1	2	1	0.58
pink	9	10	10	0.58
green	5	3	3	1.15
purple	6	6	8	1.15
blue	4	1	4	1.73
black	8	9	5	2.08
orange	2	5	7	2.52
white	10	11	6	2.65
grey	12	7	12	2.89
yellow	3	8	2	3.21
brown	11	**19**	11	4.62
magenta	**7**	21	15	7.02
other colors named by all three groups				
tan	20	14	17	3.00
gold	16	23	18	3.61
turquoise	15	**4**	**9**	5.51

The color prototypes also seem to be well established, although after the first one exemplar the agreement, revealed through frequency, dropped from 18.6 to 8.3, then to 4.8. The students were creative in their responses, for example, *as black as the shirt I am wearing; as white as a rabbit in a snowstorm; as red as the stripes on the American flag; as yellow as the National Geographic magazine; as green as the text on my shirt; as blue as the pencil I am writing with*. There were 26 different ideas for *Black*; 22 for *White*; 20 for *Red*; 24 for *Yellow*; 20 for *Green*; 17 for *Blue* (see Appendix 2). The result of *Fire* only in third place for *Red* is unusual and would seem to reflect the lack of fire in contemporary suburban life; the same way *Sheep* is low on the list for *White*. But 44 of the 58 items listed for *Red* involved fire, blood, flowers, and fruit. 44 of the *Blue* items listed involve water of some sort. 40 of the 58 items listed for *Green* involve grass or leaf vegetation. 38 of the 58 items for *Black* involve darkness, pitch, and charcoal; and for *White* involve paper, light, and snow. 37 of the 58 items listed for *Yellow* involve the sun, flowers, or fruit. This confirms the expectation that the salient features of the natural environment influence the entrenchment of color terms and prototypes.

If we compare these results with a corpus-based analysis by accessing the *Corpus of Contemporary American English* (COCA); which includes over 520 million words that have been compiled from spoken and written texts dating from 1990 to 2015, all of the principle items listed by the students in the prototype construction are found. "As black as" is found 188 times, with Night as the most frequent collocate, the same as the 12 year olds. "As white as" is found 278 times, with Snow as the most frequent collocate (18) rather than Paper. In COCA Ghost and Chalk emerge with *White*, which were not mentioned by this group of students. "As red as" is found 152 times, with the agreement of Blood as the most prototypical item. "As yellow as" is found on 23 times and results with Sun as the most frequent prototype too. "As green as" is found 87 times, "as blue as" 95 times with Grass and Sky as the principle prototypes in agreement again with the middle-school students. The Sea and Mediterranean are more frequent than Ocean and Pacific, which instead are more salient and entrenched for the Washingtonians who live right off the Pacific Ocean. COCA show Robin's egg as frequently associated with Blue, which was not mentioned by the students.

CONCLUSION

The aim of this paper has been to add an account of young English speaking middle-school students input to color listing and color prototype analysis. The two questionnaires about color terms and prototypes respond well to theories

of language processing and the prototypical color constructions. Overall, the results have shown (i) that 12 year olds in a suburban multi-ethnical society already have a well entrenched color vocabulary; (ii) at this age students are already able to use linguistic devices, such as compounding to augment the vocabulary they have at disposition; (iii) the CSI allows us to untangle the multifaceted responses arriving at an accurate picture of what terms are actually entrenched and salient for the specific group; and (iv) the construction "as [color] as x" allows the students to access the prototypes that prove to be equivalent in convention to large corpus analyses. This study should serve as a base to characterize color listing and construction queries as a strong predictor of linguistic entrenchment and salience of a specific linguistic frame in language acquisition. The theoretical claim is that the basic principles of frequency and mean position in relation to number of participants allow us to establish a relevant ranking that can be compared to other linguistic groups.

As Orians (forthcoming) states many universal human traits, such as classification of basic colors, have probably evolved in response to those stable parts of the environment. Specific behavior would be impossible in the absence of "neural filters that emphasize or de-emphasize components of aspects of information" (Marler 1961). Hence, this group of young informants is already totally aware and capable of filtering, categorizing, and selecting the responses to this type of linguistic inquiry according to their interaction with the environment. The prototypes and color terms result to be totally entrenched at this age. The theoretical base of this study stems from what Schmid calls the 'Entrenchment-and-Conventionalization Model' (cf. Schmid 2013: 106–107, 2014: 242–254; Schmid 2015; Schmid & Mantlik 2015)[3]. Thus the cognitive processes employed by this study's informants may be "subsumed" under the label *entrenchment*, whereas the social processes or agreement taking place in this speech community may be identified under the label of *conventionalization*. Schmid's model predicts that the interaction of various components makes linguistic structure emerge and be perpetuated. Linguistic entrenchment may be considered according to the salience of specific usage that evolves through use and exposure, through instruction and the development of general cognitive abilities. It is likely that this group of mostly bilingual middle-school students have a particularly high level ability to process, memorize, categorize, and respond this type of task. As Ibbotson and Tomasello specify "the problem for language acquisition is that children do not experience constructions but only utterances; they must (re-)construct for themselves the constructions of their language from the individual utterances they experience" (2009: 60); hence, more exposure, more language, more elaborate usage, greater development. It will be interesting to see if these results, degree of salience, entrenchment, and conventionalization, can be replicated for people of this age, both in other English speaking groups and in other languages.

APPENDIX 1

The 24 highest ranking colors by Cognitive Salience Index

(S = F / (N x MP)

Middle School English - mean age 12 163 colors total					**Adults English - mean age 52 126 colors total**		**Adults Italian - mean age 22 134 colors total**	
COLOR	**MP**	**FREQ**	**CSI**	**CSI RANK**	**COLOR**	**CSI**	**COLOR**	**CSI**
Red	4	28	0.2414	**1**	**Blue**	0.1922	***rosso*** (red)	0.2612
Orange	5	29	0.2000	**2**	**Red**	0.1546	***giallo*** (yellow)	0.2029
Yellow	5	28	0.1931	**3**	**Green**	0.0851	***verde*** (green)	0.1731
Blue	5	27	0.1862	**4**	Turquoise	0.0712	***blu*** (blue)	0.1531
Green	5	27	0.1862	**5**	**Orange**	0.0658	***nero*** (black)	0.1248
Purple	8	25	0.1078	**6**	**Purple**	0.0634	***bianco*** (white)	0.1189
Magenta	9	27	0.1034	**7**	**Grey**	0.0625	***arancione*** (orange)	0.1052
Black	11	26	0.0815	**8**	**Yellow**	0.0575	***viola*** (purple)	0.0977
Pink	13	29	0.0769	**9**	**Black**	0.0512	***azzurro*** (turquoise)	0.0866
White	12	25	0.0718	**10**	**Pink**	0.0494	***rosa*** (pink)	0.0860
Brown	14	25	0.0616	**11**	**White**	0.0488	***marrone*** (brown)	0.0827
Grey	14	24	0.0591	**12**	Teal	0.0413	***grigio*** (grey)	0.0787
Indigo	9	15	0.0575	**13**	Lime green	0.0392	*celeste* (light blue)	0.0637
Violet	12	19	0.0546	**14**	Tan	0.0342	*lillà* (lavender)	0.0391
Turquoise	10	13	0.0448	**15**	Lavender	0.0335	*fucsia* (fuschia)	0.0377
Gold	15	14	0.0322	**16**	Burgundy	0.0316	*indaco* (indigo)	0.0344
Silver	17	15	0.0304	**17**	Aqua	0.0300	*beige* (beige)	0.0304
Teal	13	9	0.0239	**18**	Mustard	0.0291	*oro* (gold)	0.0290
Navy blue	13	8	0.0212	**19**	**Brown**	0.0257	*porpora* (violet?)	0.0225
Tan	15	9	0.0207	**20**	Beige	0.0244	*rosso bordeaux* (burgundy)	0.0217
Light blue	14	8	0.0197	**21**	Fuchsia	0.0236	*argento* (silver)	0.0199
Dark blue	15	8	0.0184	**22**	Forest green	0.0214	*giallo ocra* (mustard)	0.0194
Fuchsia	14	7	0.0172	**23**	Gold	0.0171	*verde acqua* (aqua)	0.0183
Lavender	15	7	0.0161	**24**	Chartreuse	0.0254	*cremisi* (carmine)	0.0154

APPENDIX 2

The two responses to each color construction (as [color] as x) listed in the order of frequency. The 26 different ideas for *Black* were *Night*, *Pitch*, *Space*, *Coal*, *A screen*, *Dark*, *Cats*, *Charcoal*, *Ebony*, *A heart*, *A hole*, *Horses*, *Jet*, *Lead*, *My eyes*, *Pavement*, *Pupil*, *A raven*, *Shadow*, *Slate*, *Snow-White's hair*, *Soot*, *Tar*, *The shirt I'm wearing*, *This print*, *Witch's cat*. The 22 ideas for *White* were *Paper*, *Snow*, *Cloud(s)*, *Light*, *(no answer)*, *The sun*, *A birch tree*, *Blinding light*, *The ceiling light*, *A dove*, *Egg-whites*, *Lightning*, *My socks*, *A rabbit in snow storm*, *Sand*, *A sheep*, *The moon*, *The pants I'm wearing*, *The sclera of the eye*, *Tissues*, *A wedding gown*, *The walls.* The 20 ideas for *Red* were *Blood*, *Rose(s)*, *Fire*, *An apple*, *Cherries*, *A heart*, *Raspberries*, *(nothing)*, *The lines on the American Flag*, *The first color in rainbow*, *A Ferrari*, *A fire engine*, *An Italian Flag*, *Lipstick*, *Mars*, *My ear buds*, *Rust*, *A stop sign*, *Sun-ish*, *Tomatoes*. The 24 ideas for *Yellow* were *The sun*, *A banana*, *Dandelions*, *Lemons*, *(nothing)*, *Daffodils*, *A canary*, *The stars*, *Butter*, *Cartoon-stars*, *Corn*, *Egg-yolk*, *A field of flowers*, *Fire*, *Flowers*, *Gold*, *A bumblebee*, *A light bulb*, *The National Geographic magazine*, *Part of a flame*, *A school-bus*, *Sunflower*, *Sunshine*, *Trees in the fall*. The 20 ideas for *Green* were *Grass*, *Leaves*, *Trees*, *A Forest*, *Frogs*, *(nothing)*, *A lime*, *Emeralds*, *Wild Grass*, *The text on my shirt*, *Over brush in a tree*, *A plant*, *A pine tree*, *A high-lighter*, *Granny apples*, *A field*, *Ferns*, *Broccoli*, *Barf*, *Asparagus*, *A field*. The 17 ideas for *Blue* were *The sky*, *The ocean*, *The sea*, *Water*, *Jeans*, *A whale*, *(nothing)*, *A blueberry*, *A car*, *Eyes*, *Ice*, *My eyes*, *Painters tape*, *A pool*, *The pencil I'm writing with*, *The top left corner of the US Flag*, *US coast guard uniforms.*

ACKNOWLEDGEMENTS

I would like to thank the students at the Explorer Middle School in Everett, Washington, who kindly cooperated on this study. My special thanks go to their teacher Mary Fitzmaurice, who encourages enthusiasm to learn, creativity, and curiosity.

NOTES

[1] Construction in cognitive linguistics refers to the "chunk" of information that is stored in long-term memory in a form function couple. Ibbotson and Tomasello explain, "One of the great theoretical advances in modern linguistics is the recognition that grammatical *constructions* are conventionalized pairing of complex forms with complex semantic/pragmatic functions. Phrasal patterns and the rules of syntactic combination not only have meaning but also have the capacity to change the meanings of the words they govern" (2009: 59–60). In this study I used the construction "as [color] as X" as a structure to uncover the prototypes that are associated with the six BCTs in English. I retrieved expressions like "as black as coal", "as white as snow", "as red as blood", where the informants filled the construction gap with the object. (Cf. Construction Grammar, see Goldberg 1995.)

[2] The 163 colors named in order of the Cognitive Salience Index: Red, Orange, Yellow, Blue, Green, Purple, Magenta, Black, Pink, White, Brown, Grey, Indigo, Violet, Turquoise, Gold, Silver, Teal, Navy blue, Tan, Light blue, Dark blue, Fuchsia, Lavender, Sky blue, Neon green, Dark green, Yellow green, Crimson, Peach, Aqua, Hot-pink, Neon yellow, Baby blue, Beige, Cyan, Cerulean, Mahogany, Neon pink, Light green, Aquamarine, Scarlet, Siena, Yellow orange, Forest green, Light pink, Umber, Lime green, Blue green, Navy, Electric blue, Neon blue, Maroon, Rose, Salmon pink, Dark purple, Dark red, Chartreuse, Light red, Bronze, Dark orange, UV colors, Neon orange, Dark grey, Burgundy, Light orange, Light yellow, Off-white, Light grey, Jade, Powder blue, Rose-gold, Rust, Ultramarine, Ebony, Red orange, Cream, Periwinkle, French grey, Ocean blue, Emerald, Salmon, Cool grey, Fluorescent yellow, Lilac, *Rainbow*, Azure, Jeans, Royal blue, Carnation pink, Faded green, Hunter green, Indian yellow, Carmel, Orange-red, Russet, Sea green, Brick, Hazel, Royal violet, Smokey, Sunset, Neon purple, Blood red, Mustard yellow, Prussian blue, Butter yellow, Coral pink, Non-photo blue, *Clear*, Deep orange, Midnight blue, Rose madder, Sunset orange, Lamp black, Reddish-brown, Robin egg blue, Chinese white, Earth, Electric pink, Golden, Lemon yellow, Light indigo, Light purple, Sunshine yellow, Blue grey, Grass green, Light violet, Pastel pink, Strawberry, Yellowish green, Amber, Pastel yellow, Pea green, Neon red, Pastel blue, Pastel green, Raspberry, Mint green, Chocolate brown, Coral, Vibrant blue, Dark yellow, Bumblebee yellow, Iron, Light brown, Bright yellow, Dark indigo, Egg shell, Neon indigo, Soft pink, Orange yellow, Tickle-me-pink, Dark violet, Peach red, Gravel Grey, *Neon brown*, Red violet, *Neon black*, Violet red, Sea blue, Green-blue, Foam-green.

[3] The major elements of the model are summarized in four types of activities invariably involved in it –sensory, motor, cognitive, and social activities that constitute the core of the model, thus marking the framework as belonging to the group of usage-based models. Its key assumption is that what we generally assume to be 'language' or 'the linguistic system' comes about and is continuously updated by the interaction of two types of processes: a limited set of cognitive processes operating in the minds of speakers, subsumed under the label 'entrenchment', and a limited set of socio-pragmatic processes operating in communities, subsumed under the label 'conventionalization'. There are three entrenchment processes – association, routinization, and schematization – and four conventionalization processes – innovation, co-adaptation, diffusion, and normation. The interaction of entrenchment and conventionalization processes depends on usage and on the activities involved in usage, and it is influenced by a (probably open-ended) set of cognitive, emotive-affective, pragmatic, and social forces (Schmid & Mantlik 2015: 587).

REFERENCES

Berlin, Brent & Kay, Paul 1991 [1969]. *Basic Color Terms: Their Universality and Evolution*. Berkeley & Los Angeles & Oxford: University of California Press.

Biggam, Carole P. 1997. *Blue in Old English: An Interdisciplinary Semantic Study.* Amsterdam & Atlanta: Rodopi.

Casson, Ronald W. 1997. Color Shift: Evolution of English Color Terms from Brightness to Hue. In: C. L. Hardin & Luisa Maffi (eds.) *Color Categories in Thought and Language.* Cambridge: Cambridge University Press, pp. 224–239. http://dx.doi.org/10.1017/cbo9780511519819.010.

Corbett, Greville G. & Davies, Ian R. L. 1995. Linguistic and Behavioral Measures for Ranking Basic Colour Terms. *Studies in Language*, Vol. 19, No. 2, pp. 301–357. DOI: 10.1075/sl.19.2.02cor.

Corbett, Greville G. & Davies, Ian R. 1997. Establishing Basic Color Terms. In: C. L. Hardin & Luisa Maffi (eds.) *Color Categories in Thought and Language.* Cambridge: Cambridge University Press, pp. 197–223. http://dx.doi.org/10.1017/cbo9780511519819.009.

Croft, William & Cruse, D. Alan 2004. *Cognitive Linguistics.* Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511803864.

COCA = *Corpus of Contemporary American English*. Available at http://corpus.byu.edu/coca/, last accessed on May 24, 2016.

Davies, Ian R. L. & Corbett, Greville G. 1994. The Basic Color Terms of Russian. *Linguistics*, Vol. 32, No. 1, pp. 65–89. http://dx.doi.org/10.1515/ling.1994.32.1.65.

Davies, Ian R. L. & Corbett, Greville G. 1995. A Practical Field Method for Identifying Probable Basic Colour Terms. *Languages of the World*, Vol. 9, No. 1, pp. 25–36. Available at http://epubs.surrey.ac.uk/2090/, last accessed on May 24, 2016.

Dedrick, Don & Sayim, Bilge & Jameson, Kimberly A. & Alvarado, Nancy & Szeszel, Monika K. 2005. Color, Color Terms, Categorization, Cognition, Culture: An Afterword. *Journal of Cognition and Culture*, Vol. 5, No. 3, pp. 487–495. http://dx.doi.org/10.1163/156853705774648545.

Desgrippes, Magalie 2013. Dynamic Cognition: When the Best Example Falls Out the Boundary. *Colour Language and Colour Categorization Conference*, 4–7 June, Institute of the Estonian Language. Tallinn: Eesti Keele Sihtasutus.

Evans, Vyvyan 2007. *A Glossary of Cognitive Linguistics.* Edinburgh: Edinburgh University Press. Available at http://www.vyvevans.net/GLOSSARY.pdf, last accessed on May 24, 2016.

Geeraerts, Dirk. 2009. Cognitive Semantics. In: *Theories of Lexical Semantics.* Oxford: Oxford University Press, pp. 182–266. DOI:10.1093/acprof:oso/9780198700302.001.0001.

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago & London: University of Chicago Press.

Hardin, C. L. & Maffi, Luisa (eds.) 1997. Introduction. In: C. L. Hardin & Luisa Maffi (eds.) *Colour Categories in Thought and Language.* Cambridge: Cambridge University Press, pp. 1–18. http://dx.doi.org/10.1017/cbo9780511519819.001.

Ibbotson, Paul & Tomasello, Michael 2009. Prototype Constructions in Early Language Acquisition. *Language and Cognition*, Vol. 1, No. 1, pp. 59–85. http://dx.doi.org/10.1515/LANGCOG.2009.004.

Kay, Paul & Berlin, Brent & Maffi, Luisa & Merrifield, William R. & Cook, Richard S. 2009. *The World Color Survey.* Stanford: CSLI.

Kuehnast, Milena & Wagner, Valentin & Wassiliwizky, Eugen & Jacobsen, Thomas & Menninghaus, Winfried 2014. Being Moved: Linguistic Representation and Conceptual Structure. *Frontiers in Psychology*, Vol. 5: 1242. http://dx.doi.org/10.3389/fpsyg.2014.01242.

Lakoff, George 1990 [1987]. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* Chicago & London: The University of Chicago Press.

Langacker, Ronald W. 2008. *Cognitive Grammar: A Basic Introduction.* Oxford & New York: Oxford University Press. doi:10.1093/acprof:oso/9780195331967.001.0001.

Marler, Peter 1961. The Logical Analysis of Animal Communication. *Journal of Theoretical Biology*, Vol. 1, No. 3, pp. 295–317. http://dx.doi.org/10.1016/0022-5193(61)90032-7.

Orians, Gordon H. forthcoming. Our Biological Past in Our Modern Verbal World. In: Annalisa Baicchi & Remì Digonnet & Jodi L. Sandford (eds.) *Epistemology, Embodiment, and Language: Sensory Perceptions and Representations.* Studies in Applied Philosophy, Epistemology and Rational Ethics, Berlin: Springer.

Pitchford, Nicola J. 2006. Reflections on How Color Term Acquisition is Constrained. *Journal of Experimental Child Psychology*, Vol. 94, No. 4, pp. 328–333. http://dx.doi.org/10.1016/j.jecp.2006.02.004.

Pitchford, Nicola J. & Kathy T. Mullen, 2002. Is the Acquisition of Basic-Color Terms in Young Children Constrained? *Perception*, Vol. 31, No. 11, pp. 1349–1370. DOI:10.1068/p3405.

Rosch, Eleanor 1975. Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology: General*, Vol. 104, No. 3, pp. 192–233. http://dx.doi.org/10.1037/0096-3445.104.3.192.

Rosch, Eleanor 1978. Principles of Categorization. In: Eleanor Rosch & Barbara B. Lloyd (eds.) *Cognition and Categorization.* Hillsdale, N.J.: Laurence Erlbaum Associates, pp. 27–48. Available at http://commonweb.unifr.ch/artsdean/pub/gestens/f/as/files/4610/9778_083247.pdf, last accessed on May 25, 2016.

Rosch, Eleanor 1983. Prototype Classification and Logical Classification: The Two Systems. In: Ellin Kofsky Scholnick (ed.) *New Trends in Conceptual Representation: Challenges to Piaget's Theory?* Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 73–86.

Sandford, Jodi L. 2012. *Red Clover. Linguaggio e percezione dei colori: uno studio cognitivo applicato all' inglese*. Roma: ARACNE editrice S.r.l.

Sandford, Jodi L. 2015. A Cognitive Linguistic Usage Perspective: What is Italian Blu, azzurro, celeste? Do English agree on BLUE semantics? *Cultura e Scienza del Colore / Color, Culture, and Science*, Vol. 4, pp. 22–30. Available at http://jcolore.gruppodelcolore.it/numeri/pdf/R0415_v04_11_2015_secured.pdf, last accessed on May 25, 2016.

Sandford, Jodi L. forthcoming. Cognitive Entrenchment of Color Category Associates and Implicit Attitudes in English. In: Mari Uusküla & Geda Paulsen & Jonathan Brindle (eds.) *Colour Language and Colour Categorization*. Newcastle upon Tyne: Cambridge Scholar Publishing.

Schmid, Hans Jörg 2007. Entrenchment, Salience, and Basic Levels. In: Dirk Geeraerts & Hubert Cuyckens (eds.) *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford University Press, pp. 117–138. DOI:10.1093/oxfordhb/9780199738632.013.0005.

Schmid, Hans Jörg 2013. Is Usage more than Usage After All? The Case of English *not that*. *Linguistics*, Vol. 51, No. 1, pp. 75–116. http://dx.doi.org/10.1515/ling-2013-0003.

Schmid, Hans Jörg 2014. A Framework for Understanding Linguistic Entrenchment and Its Psychological Foundations in Memory and Automatization. In: Hans Jörg Schmid (ed.) *Entrenchment, Memory and Automaticity: The Psychology of Linguistic Knowledge and Language Learning*. Berlin: De Gruyter Mouton, pp. 1–24. Available at https://www.academia.edu/8953816/A_framework_for_understanding_linguistic_entrenchment_and_its_psychological_foundations_in_memory_and_automatization._Download_from_homepage_, last accessed on May 25, 2016.

Schmid, Hans Jörg 2015. A Blueprint of the Entrenchment-and-Conventionalization Model. *Yearbook of the German Cognitive Linguistic Association*, Vol. 3, No. 1, pp. 3–26. DOI: 10.1515/gcla-2015-0002.

Schmid, Hans Jörg & Mantlik, Anette 2015. Entrenchment in Historical Corpora? Reconstructing Dead Authors' Minds from their Usage Profiles. *Anglia*, Vol. 133, No. 4, pp. 583–623. http://dx.doi.org/10.1515/ang-2015-0056.

Sutrop, Urmas 2000. The Basic Colour Terms of Estonian. *Trames*, Vol. 4, No. 2, pp. 143–168.

Sutrop, Urmas 2001. List Task and Cognitive Salience Index. *Field Methods*, Vol. 13, No. 3, pp. 263–276. http://dx.doi.org/10.1177/1525822X0101300303.

Sutrop, Urmas 2002. *The Vocabulary of Sense Perception in Estonian: Structure and History*. Frankfurt am Main & Berlin & Bern & Bruxelles & New York & Oxford & Wien: Peter Lang.

Talmy, Leonard 2000. *Toward a Cognitive Semantics. Volume II: Typology and Process in Concept Structuring*. Cambridge, MA: MIT Press. Available at http://linguistics.buffalo.edu/people/faculty/talmy/talmyweb/TCS.html, last accessed on May 25, 2016.

Talmy, Leonard 2005. Foreword. In: Monica Gonzalez-Marquez & Irene Mittelberg & Seana Coulson & Michael J. Spivey (eds.) *Methods in Cognitive Linguistics*. Amsterdam & Philadelphia: John Benjamins Publishing Company, pp. xi–xxi. DOI: 10.1075/hcp.18.

Uusküla, Mari & Sutrop, Urmas 2007. Preliminary Study of Basic Colour Terms in Modern Hungarian. *Linguistica Uralica*, Vol. 43, No. 2, pp. 102–123. Available at http://www.eap.ee/public/Linguistica_Uralica/2007/issue_2/ling-2007-2-4.pdf, last accessed on May 25, 2016.

Uusküla, Mari 2007. The Basic Colour Terms of Finnish. *SKY Journal of Linguistics*, Vol. 20, pp. 367–397. Available at https://www.academia.edu/9109314/The_basic_colour_terms_of_Finnish, last accessed on May 25, 2016.

Uusküla, Mari 2008. The Basic Colour Terms of Czech. *Trames*, Vol. 12, No. 1, pp. 3–28. DOI: 10.3176/tr.2008.1.01.

http://dx.doi.org/10.7592/FEJF2016.64.glukhova

COLOUR SYMBOLS IN MARI SONGS

Natalia Glukhova

Abstract: This article presents a system of colour symbols in Mari folk songs based on the results of a multifold investigation. The research was carried out with the help of a complex technique applied to 2100 songs from different song collections. Mari colour symbols have never before been the object of research. The process of investigation included several steps. The most important of these was the semantic analysis that helped to discern 2000 mentions of four main spectre colours as well as white, black, silver, and golden. Quantitative data evaluation singled out a dominant group by a dichotomous method, applying the principle of simple majority employed in mathematical statistics. The same technique divided the other colour symbols into complementary, auxiliary, and insignificant groups. The results of an investigation into ethnic symbology are also shown graphically. The main reconstructed meanings of colour in the analysed songs denote such emotions as joy, wonder, astonishment, grief, melancholy, some aesthetic ideals, ethical vices, as well as people's character and appearance.

Keywords: auxiliary, complementary, complex technique, dominant, folk songs, insignificant groups, ranking/rating, semantic analysis, symbol, systems theory

INTRODUCTION

Symbol

Being one of the fundamental concepts of philosophy, aesthetics, and arts, symbol has a long history of research, interpretation, and comprehension. Among works dedicated to symbol studies Aleksei Losev's monographs and articles were the source of inspiration for home scholars and were considered to be of utmost importance for symbology. In one of his monographs the author offers a logical analysis of nine moments that constitute symbol, as well as classification of symbol types to which he refers as scientific, philosophical, mythological, religious, ideological, artistic, determinative, and technical symbols (Losev 1995 [1976]: 264–274).

Symbol is compared to such art categories as 'sign', 'allegory', 'personification', 'artistic image', 'type', 'metaphor', and 'myth'.

Yuri Lotman in his works on symbol shows that symbol represents a certain text both on the plane of expression and on the plane of content, i.e. has

a single meaning and a distinctly expressed border allowing distinguishing it from a surrounding semiotic context (Lotman 1992: 192). In this case a symbol is a kind of semiotic condenser, playing the role of a mediator between the synchrony of the text and cultural memory (Lotman 2000: 249). Yuri Lotman's valuable observations on symbols are also concerned with its archaic character and the fact that symbols represent one of the most stable elements of the cultural continuum.

Generalisation of different theories on the correlation of sign and symbol, the interpretation of symbol in art, literature studies, philosophy, and in different schools of semiotics is presented in the publications of the last decade. Thus, the history of a theoretical comprehension of symbol in philosophy, aesthetics, philology, and psychoanalysis is described in Sergei Averintsev's works, where the author pays much attention to the understanding of symbol in different schools of thought in the 20th century (Averintsev 2001: 155–161).

A concise overview of foreign scholars' theories on the nature of symbols used in various cultures, rituals, and everyday life is offered in the article written by Edit Turner in *Folklore: Encyclopedia of Beliefs, Customs, Tales, Music, and Art* (Turner 1997: 25–29). She briefly mentions Émile Durkheim's works and his school in France at the beginning of the 20th century, discusses Edward Sapir's works who showed the complexity of symbol as a multivalent phenomenon having many meanings, and points at Victor Turner's research. She also discusses the understanding of symbols by Sigmund Freud, Carl Gustav Jung, Raymond Firth, Claude Lévi-Strauss, Lindsey K. Girtz, and Don Handelman (Turner 1997: 28).

Philologists studying the folklore of Finno-Ugric peoples usually refer to the works of Russian scholars who studied traditions and customs of Slavic ethnoses. They are Aleksandr Veselovskii, Aleksandr Potebnia, Boris Sokolov, and Yuri Sokolov. In their investigations there are explanations of psychological origin of peoples' symbolics (Veselovskii 1940: 107–117), the meanings of separate symbolic images in folk poetry (Potebnia 1989: 285–378), as well as a social aspect of symbols in people's lyrics (Sokolov 1941: 125–456).

Taking into consideration the works on symbol I use a working definition in this article. Symbol is an object/phenomenon of extra-linguistic reality denoted by a lexeme which is the first member of a two-member parallel construction in the song without a second part of the parallel. Symbol is regarded, considering its denomination, the number of its meanings, and the probability of its usage that testifies to its importance in the ethnic folk song genre.

Colour symbolism

Universal symbols among all nations are names of the colours of natural objects. Scholars, artists, poets, composers, and writers have provided various classifications of colours, and investigated colour meanings as well as their emotional impact on people. There is extensive literature on the subject, which can be recommended to the readers (see: Vezhbitskaia 1996: 231–291; Goethe 1957: 300–340; Deribere 1964; Frumkina 1984; Yanshin 2006: 18–91; Berlin & Kay 1991 [1969]; Uusküla & Sutrop 2007: 102–123; Uusküla 2007: 367–397; Davies & Corbett 1994: 65–89; Uliashev 1999; Ryabina 2011).

In the analysed material of 2100 Mari folk songs there are 2000 mentions of four main colours of the spectre, as well as white and black. Dominant colours are *white*, *black,* and *silver*; they constitute 59.3% of all colour mentions in the investigated songs. *Yellow*, *red*, and *green* compose the next group by their frequency of mention, and *blue*, *golden*, *brown*, and *grey* complete the colour palette of the analysed Mari songs.

In connection with this fact it is quite appropriate to cite O. Spengler's words on colour symbols of different ethnoses, which testify to some coincidence of semantics in Mari songs with the interpretation of colours in ancient cultures. He argued that *yellow* and *red* are ancient paints of matter, closeness, popular colours, colours of people's masses. *Blue* and *green* are colours of loneliness and solitude, care, ties of a present moment with the past and future, destiny, colours of fate (Spengler 2000: 360–374).

In P. Yanshin's *Psikhosemantika tsveta* (Psychosemantics of colour) there is a part on the evolution of colour thesaurus and enrichment of colour archetype in different languages and cultures. Referring to Berlin and Kay's work (1991 [1969]) the author points out the presence of primordial simplest colour vocabulary containing only two designations – for *black* (dark) and *white* (light). Later on *red* is added to *black* and *white*, and in this case *yellow*, *orange,* and *brown* might be included in the group of 'red'. *White*, *black*, and *red* are considered the earliest colour designations among many peoples. *Blue*, *yellow,* and *green* followed them (Yanshin 2006: 162–170).

The combination of colours and their meanings in the cultures of Finno-Ugric peoples coincides in some way and differs from semantics described on the materials of other ethnoses. This will be demonstrated on the examples of Mari songs.

METHODOLOGICAL FRAMEWORK

The research combines methods and techniques of semantic analysis to which I refer componential and contextual types of analyses and systems theory. Systems theory, in turn, involves factor and statistical analysis, including the principle of a simple majority of a dichotomous method. Thus the investigation into colour symbolism consists of several stages worked out by Vladimir Glukhov, which were applied to different Mari folklore genres.

The algorithm of the research includes the following steps:

1. The process of reading texts with the aim of discerning symbols (referred to with the unifying term 'factor') after applying the componential and contextual types of text and lexeme analysis;
2. The compilation of the list of factors (symbols);
3. The distribution of the texts according to the factors;
4. The estimation of factor incidence and calculation of the probability of factor usage (the preparation of tables);
5. The ranking of factors in the descending order of probability (the preparation of probability distribution diagrams);
6. The singling out of dominant, complementary, auxiliary, and insignificant factors by a dichotomous method, applying the principle of simple majority employed in mathematical statistics (Glukhov & Glukhova 2013: 399–412).

Thus at the beginning of the research texts of songs were read and analysed with the aim of determining symbols. Psychological parallelism described by folklore researchers helps to discern symbols in the investigated songs. According to their findings symbol is a lexeme that is the first member of a two-member parallel construction without a second part of the parallel.

As was already mentioned, in this article symbol is also viewed with the number of its meanings and the frequency of its mentioning taken into account. The more important a symbol is, the more often it is mentioned in the songs. This fact allows the ranking of symbols in a descending order of their occurrence. The symbol incidence has been esteemed and the probability of symbol usage calculated. Then the data have been further presented in a tabulated form. Furthermore the numerical data to which a dichotomous method is applied help to discern four groups of symbols. They are conventionally divided into *dominant*, *complementary*, *auxiliary*, and *insignificant*.

The next stage was devoted to the preparation of diagrams with the ranking of symbols in a descending order of probability. In the summarising diagram each symbol is shown as a separate column with a height proportional to its

probability of occurrence which is shown in the tables. The columns are arranged in descending order.

There are other approaches and methods of colour analysis in different languages (Davies & Corbett 1995: 25–36; Corbett & Davies 1995: 301–357.) But to reveal a list of colour symbols and their semantics, I resorted to my own methods and techniques which were prompted by the material itself.

The investigation into traditional Mari symbolism in folk songs concludes with a summarisation of the results.

RESULTS AND DISCUSSION

Specific survey findings show that folk songs occupy a particular place in Mari folklore. There are special studies, monographs, and articles dedicated to the classifications of Mari songs, specific features of texts, and musical accompaniment belonging to different sub-ethnic groups. An analytical survey of the ideas is represented in the monograph titled *Sistemy simvolov pesen finno-ugrov Uralo-Povolzh'ia* (A System of Symbols in the Songs of Volga-Ural Finno-Ugrians) (Glukhov & Glukhova 2012: 44–48).

Mari songs are divided into two groups: songs connected with different ceremonies and rites (ritual songs) and non-ritual songs. Ritual songs are closely connected with different calendar holidays and family celebrations (wedding songs, remembering songs; lullabies). Non-ritual songs are also numerous. They are 'working' songs, guest songs, orphan's songs, recruit songs, comic songs, love songs, and meditation (contemplation) songs.

The analysis of the subsystem of symbols has been carried out on the material of Mari proper names and folk songs. The article presents the results of investigation of such songs as wedding songs, remembering songs, recruit songs, guest's songs, love songs (mainly), comic songs, and meditation songs without pointing at the genre as the main idea was to show a system of symbols in Mari songs despite their type.

The analysis has shown that in Mari ethnic culture, emotions and feelings are expressed with the help of *five leading groups of symbols.* Statistical analysis, together with a dichotomous method, reveals the most widespread names of *plants*, *numerals*, *colours*, *birds*, and *animals*, organised into groups of *complementary*, *auxiliary*, and *insignificant* symbols, typical of Mari folk songs. The data on each group have been shown in tables and then presented in the form of probability diagrams of names of *plants*, *colours*, *numerals*, *birds*, and *animals* mentioned by number in Mari songs.

Mari system of symbols in songs

The multifold investigation applied to 2100 songs from different collections of Mari folk songs (Mari kalyk muro 1951; Podelkov 1976; Kulshetov 1990; Vasilyev 1991; Ervel marii muro-vlak 1994; Pesni gornykh mari 2005) allowed to compile a peculiar ranking of symbols mentioned in the analysed texts – names of flora, object colours, numerals, and fauna (Table 1).

Table 1. *Ranking of symbols mentioned in Mari folksongs.*

No.	Type of symbol	Amount (number of mentions)	Probability
1	Plants, trees, flowers, berries	2966	0.432
2	Colours of objects	1554	0.226
3	Numbers	882	0.128
4	Birds	836	0.122
5	Animals	630	0.092
Total sums		**6868**	**1.000**

The data in the table have been used for drawing a diagram of the probability distribution of symbols mentioned in Mari folk songs. Colour symbols occupy the second place.

There is no specific research into colour symbolism in Mari songs. This is the first attempt to show the system of colours and their meanings in Mari culture.

The analysed texts present 2000 mentions of four main spectre colours as well as *white*, *black*, *silver*, and *golden*.

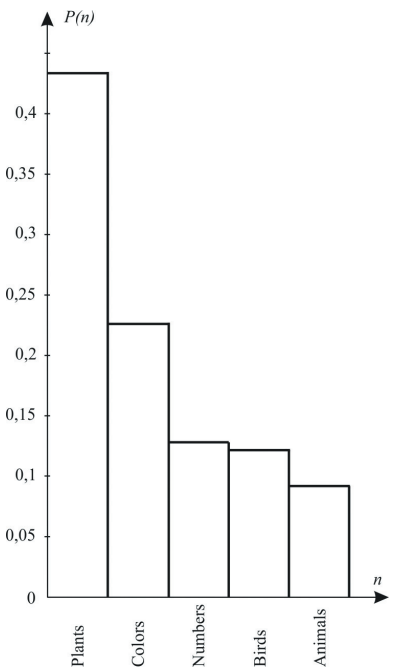

Figure 1. *Probability distribution of symbols mentioned in Mari folk songs.*

Statistically, names of colours are distributed in the frequency of their mention and shown in Table 2.

Table 2. *Results of the analysis of the number of colour mentions in Mari folk songs.*

No.	Colour	Number of mentions	Probability of usage	No.	Colour	Number of mentions	Probability of usage
1	White	448	0.288	6	Green	127	0.082
2	Black	276	0.178	7	Blue	83	0.053
3	Silver	198	0.127	8	Golden	66	0.042
4	Yellow	184	0.118	9	Brown	33	0.021
5	Red, pink, scarlet	127	0.082	10	Grey	12	0.008
				Sums		**1554**	**1.000**

The table shows that dominant colours in the research material are: *osh, osho* (white), *shem*, *sheme* (black) and *shii*‘ (silver), as they comprise 59.3% of all mentions in the investigated songs. *Narynche* (yellow), *yoshkar*, *yoshkarge* (red and its hues), and *užarge* (green) constitute the next group calculated with the method of consecutive dichotomy (28.2%). The third and the fourth groups include four colours: *kande* (blue), *shörtn’**ö*** (golden) (12.5%), *küren* (brown), and *sur* (grey) (3%). Figure 2 presents a diagram drawn on the basis of the data in the table.

A semantic analysis of the songs revealed the meaning of ten colours. The colour names function as epithets in the analysed songs. However, colour characteristics themselves can have their own meanings, which can be seen in the enumerated examples presented in the article.

Here mainly the examples from the dominant group are considered. The most polysemantic colour is *white*. In Mari old songs there are tautological epithets with this lexeme. It can be seen in the following example:

Pasu gyna pokshelne kugu tumo,	In the middle of the field there is a big oak,
Kugu tumo jymalne – osh kue.	Under the big oak there is a white birch.
Kugu tumo, achaem, tin’ ulat,	The big oak is my father,
Osh kuežym parchažym min’ ulam	And the branches of white birch are me.
(Ervel marii muro-vlak 1994: 68)	

The adjective *osho* (white) is associated with a happy life in a big family, together with all relatives (Ervel marii muro-vlak 1994: 23, 61, 67). The word ‘white’ also expresses the ideas of cleanliness and neatness in the house (Ervel marii muro-vlak 1994: 225).

The adjective 'white' as an attribute to young girls' clothes shows the naivety of youth and hope of happiness. And yet, life can pass by without realising the cherished dreams.

Osh shavyr chien, parsyn savyc jalshten, Järmingäsh kedeok, ÿdÿr kurymem ertÿsh. …	In a white linen cloth, in a silk kerchief, I did not go to a market place, but the girlhood slipped by.
Osh tygyr chien, jažo zapon pižÿkten, Shÿnzÿshvolkysh läkdeok, ÿdÿr kurymem ertÿs (Pesni gornykh mari 2005: 56–57).	In a white dress, in a beautiful apron I have never been to the village gatherings, but the girlhood slipped by.

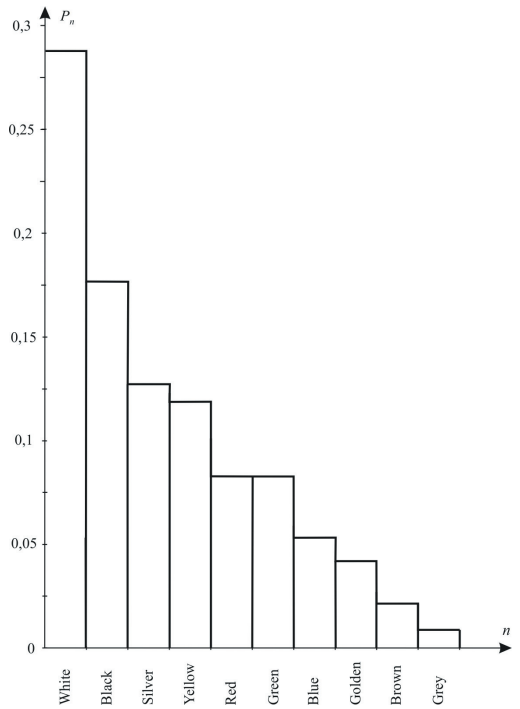

Figure 2. *Probability distribution of colours mentioned in Mari songs.*

The white colour of a girl's kerchief is associated with her feelings for a young man:

Osh shovychym osho gynat,	Though a white kerchief is white
Ere mushmem vele shuesh.	I always want to wash it.
Ala, joltashem, syrenat gynat,	Though my beloved is cross (with me),
Ere užmem vele shuesh	I always want to look at him.
(Mari kalyk muro 1951: 151).	

In the song *Have you climbed up the hill?* there is colour antithesis – white *vs.* black. If *white* was a symbol of love, the colour *black* denotes the end of romantic feelings, parting with the young man:

Arka ümbak küzyshych mo?	Have you climbed up the hill?
Shem pyl kaimym užych mo?	Have you seen a black cloud flying by?
Oj, joltashem, syryshych mo?	Oh, my friend, aren't you angry with me?
Vese dene kajymem užych mo?	Have you seen that I was walking with another guy?
Myjže pushym osh perchatkym,	I have given you the white gloves
Tyjže pushych shememden	You have returned me the black ones.
Myjže modym chon jöralten	I loved you with all my heart
Tyjže modych kudaltash shonen	And you thought to leave me.
(Mari kalyk muro 1951: 137–138).	

White can express sorrow of the singer coinciding with the meaning of *black*:

Osh savycym jalshtyshym	I put on a white kerchief,
Ÿshkežÿ oshy ylamat.	And I was pale myself.
Licäem, oshym, pälevÿ,	Everybody looked at my pale face,
Mÿn'ÿn oixem ÿsh pälen.	But nobody learned about my sorrow.
(Pesni gornyh mari 2005: 70).	

Black is associated with sorrow, and yearning for a beloved girl:

Jua voktene olykyshtet shüshken	Near the River Ya, on the meadow
Koshtsho shüshpyk min' yl'ym,	I was a singing nightingale.
Shem kožlashke purymem godym	When I entered a dark (lit. *black*) forest,
Imne kajysh vüd vokten,	My horse went along the river.
Ej! Sagynal'ym tin'ym oilen	Oh! And I remembered you with yearning.
(Vasilyev 1991: 89).	

The colour black symbolises a strange land and can also be a symbol of danger in communication with distant relatives:

Shem kožla kornyžym legyldalash	To pass through a black forest you need
Shishte dech sare-laj imne külesh.	a light-brown horse, yellower than wax.
Ty gyna-laj shochshem den ojlasalash	To talk with this relative
Porsyn dechat pushydi mut külesh	you need words softer than silk.
(Ervel marii muro-vlak 1994: 20).	

Black used to describe girls' eyes is compared to black currants and testifies to their sexual maturity and readiness for marriage.

Kuryk kydalne joshkar isnege	There are red berries on the hillside;
Kuryk laj vujyshto shemshoptyr	There is blackcurrant on the hillside.
Shemshoptyr gane shem shinchat dene	Flirting with your eyes like
Modyktltyn, memnam savyret.	blackcurrant, you are seducing us.
(Ervel marii muro-vlak 1994: 94).	

In the following excerpt the idea is the same as in the previous text:

Flashing our black eyes	First we simply give him shelter and
We give shelter to a handsome man	then marry him (Podelkov 1976: 98).

Black colour may also characterise beauty of things, clothes, and beloved girl's hair.

Silver is defined as third in frequency in the analysed material. The word *shii* (silver) in combination with nouns may have the meaning of 'beautiful', 'beloved', 'dear':

Lijdalalže yle shij sambarem,	Let the silver samovar be.
Shinchalalže yle üstembalnem.	Let it stand on my table.
Lijdalalže yle yshkemyn joltashem,	Let me have a dear friend.
Shogalalže yle pelenem.	Let him always be with me
(Kulshetov 1990: 124).	

In some examples, the word *silver* defining jewellery (rings, earrings, beads) has a direct meaning pointing to a metal of which the jewellery was made, thus emphasising the material prosperity of a person (Ervel marii muro-vlak 1994: 186) or a lost youth.

In the song *Aidyza lai udyrlakem* (Let us, girls, go out) the combination of the word 'silver' with the noun 'fish' symbolises *young girls*, *fish play* – girls' coquettish behaviour.

Ajdyza laj üdyrlakem	Let us, girlfriends, go out to
Ajdyza laj üdyrlakem	the river bank.
Vüd türyshkö kaena	Let us watch
Shij kolyn igyžyn	silver fish playing.
Oj, modmyžym onchena.	Oh, let us watch them playing!
Ajdyza laj üdyrlakem	Let us, girlfriends,
Küshkö ülych küzena.	walk up the street and
Küshyl muchashyn üdyrlakyn	watch the girls' habits and behaviour there.
Oj, kojyshyshtym onchena.	Oh, let us watch them.
(Mari kalyk muro 1951: 235)	

The adjectives 'red', 'pink', 'scarlet' are used in the investigated songs to show joy:

Ojgyren-ojgyrenat, tüsym pytysh,	I have become pale with grief,
Ajsta shürgesh al chijam jygena.	Let us paint our faces with pink colour
Shochshyna-laj ukelan jalt jam uke,	(use pink make-up)
Ajsta törzash al säs'käm shyndena.	We have no relatives, thus we have no joy.
(Ervel marii muro-vlak 1994: 133).	Let us plant pink flowers on the
	window sills.

These colours may 'correlate' with youth and beauty (Ervel marii muro-vlak 1994: 62). *Pink* may be a symbol of mother (ibid.: 281). *Yellow* and *blue* with the combination of field flower names are associated with the transiency of youth, beauty of the beloved friend:

Sadvichesh peledesh klovoj vujan peledÿsh,	There are blue cornflowers in the garden.
Klovoj vujan peledÿsh – mloec kurym.	A blue cornflower is a boy's life-time.
Alyk lapesh peledesh narynzy vujan peledÿsh,	There are yellow flowers in the fields.
Narynzy vujan peledÿsh – ÿdÿr kurym.	A yellow flower is a girl's life-time.
(Pesni gornyh mari 2005: 70–71).	

Yellow symbolises melancholy, a yearning to see relatives, grief that leads to early ageing. However, *a yellow flower*, *a yellow ribbon*, *a yellow kerchief* may show the beauty of a beloved girl and relatives.

Memnan urem ülykö tajyl	Our street is sloping down
Pördyn kaja sar olma.	A yellow apple is rolling easily.
Tenii kuchymo javyl tangem –	My best friend chosen recently –
Sajlen nalme sar olma	Is the best yellow apple of all.
(Vasiliev 1991: 19).	

In many Hill Mari songs the *yellow* colour of a ribbon, kerchief, apron, or shirt symbolises beauty.

Melancholy of a person because of parting with relatives or close friends is shown by the word *blue* in the analysed songs. The death of a relative is symbolised by *a blue flower*. Getting old causes sorrow and melancholy; regret regarding the lost youth is also shown with the help of *a blue flower*. The colour might mean *deceit*, and it is also associated with orphanhood. In the analysed texts *youth* and *immaturity* are expressed by the adjective 'green':

Užar gyna lepen'yn shuldyržo dene	Who saw clothes made of green
Vurgem yshten chijalshe ulo gyn?	butterfly's wings?
Užar gyna vuemym, užar kapemym	Who thinks about me, being so young
Ajdemeshat shotlysho ulo gyn?	and inexperienced [lit.: *with my green*
(Ervel marii muro-vlak 1994: 10).	*head and green body*] seriously?

But the word combination 'green-blue' symbolises the beauty of the world.

The word *golden* may be used both in direct (description of jewellery) and indirect meanings (description of wealth). It may have the meaning of 'dear' when it concerns parents or native land. It also has the meaning of 'beautiful', 'good'.

A grey duck is a symbol of a beloved girl. To have *grey eyes* is not very good in a singer's opinion. One should be dark-eyed. When a girl grows up, one should not "look at her with grey eyes": the context shows the meaning of 'with envy' or 'with dislike'. In songs for children the word *sur* (grey) is used to describe a mouse or a cat.

CONCLUSION

The history of the study of symbol is very long. Symbol was compared to such literature and arts categories as sign, allegory, personification, artistic image, metaphor, and myth. Scholars write on the correlation of sign and symbol in arts, literature studies, and philosophy, giving different definitions in different schools of semiotics. Symbols, being archaic in their nature, are the most stable elements of culture, which ensures a long-standing interest in their structure and semantics.

In Mari songs I defined five leading subsystems of symbols. With the help of the rating of subsystems the most frequent group of symbols was found. Colour symbols occupy the second place in the hierarchy of symbols in the analysed songs. Colour symbolism in different cultures as well as the nature of a symbol has a long-lasting history of studies. Nevertheless, colour symbolism on the material of Finno-Ugrians' songs, Mari in particular, has never been the subject of investigation, especially with the described methods and techniques. The conducted research has shown that the triad of dominant colour symbols in Mari song culture embraces *white*, *black*, and *silver*. Their semantics, on the one hand, coincides with the meanings of the objects' colour among other ethnic groups, and, on the other hand, is characterised by certain differences, which are vividly shown by the examples from the analysed songs.

As a result of the analysis of the texts, I have compiled an integrated table, which includes colours of auxiliary and insignificant groups. The table vividly demonstrates the multiple meaning of words – colour denominators in the researched material.

Table 3. *Colour symbolism in Mari folk songs.*

Symbol	**Semantics**
The dominant group of symbols	
White	Life with relatives, happy life in the family, cleanliness in the house; naivety, hopes of youth, first love; sorrow; wealth; beauty; respect
Black	Grief; melancholy, foreign country, strange land; danger; beauty of things, beauty of the beloved person
Black currant (description of eyes)	Readiness of young girls to get married
Silver	Beloved friend; beauty; wealth (possession of silver jewellery)
The complementary group of symbols	
Red, pink, scarlet	Joy, youth, beauty
Yellow	Melancholy, desire to see relatives; beauty; joy
Green	Youth; immaturity
The auxiliary and insignificant groups of symbols	
Blue	Yearning for relatives; melancholy about the gone youth; death; deceit
Golden	Wealth (possession of golden jewellery); beloved, favourite (about parents, relatives, native land); beautiful; good
Brown	Beauty (of things)
Grey	Ordinary, not bright, inconspicuous

The reconstructed meanings of colour in the analysed Mari folk songs mainly denote emotions (joy, wonder, astonishment, grief, melancholy); some aesthetic ideals (beauty of the beloved); vices (deceit, lies); people's character and appearance. It is interesting to note that *white* and *black* may combine with opposite meanings. Along with the description of happiness, *white* can express coldness in people's relations, discomfort, heavy feelings, and grief. The meaning of *black* contains both negative connotations that were shown in the examples and the ideal of beauty, bright colours of youth.

REFERENCES

Averintsev, Sergei 2001. Simvol. [Symbol.] In: *Sofiia-Logos: Slovar'*. [Sofia-Logos: Slovar'.] Kiev: Duh i Litera, pp. 155–161. Available at http://ec-dejavu.ru/s-2/Symbol.html, last accessed on May 18, 2016.

Berlin, Brent & Kay, Paul 1991 [1969]. *Basic Color Terms: Their Universality and Evolution*. Berkeley & Los Angeles & Oxford: University of California Press.

Corbett, Greville G. & Davies, Ian R. L. 1995. Linguistic and Behavioural Measures for Ranking Basic Colour Terms. *Studies in Language*, Vol. 19, No. 2, pp. 301–357. Available at http://www.ingentaconnect.com/content/jbp/sl/1995/00000019/00000002;jsessionid=2ckmtpvp5f276.alice, last accessed on May 18, 2016.

Davies, Ian R. & Corbett, Greville G. 1994. The Basic Color Terms in Russian. *Linguistics*, Vol. 32, No. 1, pp. 65–89. http://dx.doi.org/10.1515/ling.1994.32.1.65.

Davies, Ian R. & Corbett, Greville G. 1995. A Practical Field Method for Identifying Probable Basic Colour Terms. *Languages of the World*, Vol. 9, No. 1, pp. 25–36. Available at http://epubs.surrey.ac.uk/2090/, last accessed on May 18, 2016.

Deribere, Moris 1964. *Tsvet v deiatel'nosti cheloveka*. [Colour in Man's Activity.] Translated edition, abridged from French by A. M. Grigoryev.] Moscow: Stroiizdat.

Ervel marii muro-vlak 1994 = Gadiatov, G. (comp.) *Ervel marii muro-vlak. Pesni vostochnyh mari*. [Songs of the Eastern Maris.] Yoshkar-Ola: Mariiskoye knizhnoye izdatel'stvo.

Frumkina, Revekka 1984. *Tsvet, smysl, skhodstvo*. [Colour, Meaning, Similarity.] Moscow: Nauka.

Glukhov, Vladimir & Glukhova, Natalia 2012. *Sistemy simvolov pesen finno-ugrov Uralo-Povolzh'ia*. [A System of Symbols in the Songs of Volga-Ural Finno-Ugrians.] Moscow: Spetskniga.

Glukhov, Vladimir & Glukhova, Natalia 2013. A System of Techniques and Stratagems for Outlining a Traditional Ethnic Identity. In: Frog & Pauliina Latvala (eds.) *Approaching Methodology*. Second revised edition with an introduction by Ulrika Wolf-Knuts. Helsinki: Finnish Academy of Science and Letters & the Finnish Society of Science and Letters, pp. 399–412.

Goethe 1957 = Gete, Iogann. K ucheniiu o tsvete: Khromatika. [On Studying Colour: Chromatics.] *Izbrannye sochineniia po estestvoznaniiu*. [Selected Works on Natural Science.] Moskva & Leningrad: Izdatel'stvo AN SSSR, pp. 300–340.

Kulshetov, D. M. 1990. *Shernur kornet: Marii kalyk muro-vlak.* [A Sernur Road: Mari Folk Songs.] Yoshkar-Ola: Respublikanskii metodicheskii tsentr kul'tury MASSR.

Losev, Aleksei 1995 [1976]. *Problema simvola i realisticheskoe iskusstvo*. [The Problem of Symbol and Realistic Art.] Moscow: Iskusstvo. Available at http://www.gumer.info/bibliotek_Buks/Culture/Los_PrSimv/index.php, last accessed on May 18, 2016.

Lotman, Yuri 1992. Simvol v sisteme kul'tury. [[Symbol in a System of Culture.] *Izbrannyie stat'i v tryoh tomah. Tom I: Stat'i po semiotike i topologii kul'tury*. [Selected Articles in Three Volumes. Volume I: Articles on Semiotics and Topology of Culture.] Tallinn: Aleksandra, pp. 191–199.

Lotman, Yuri 2000. *Semiosfera.* [Semiosphere.] Sankt Peterburg: Iskusstvo. Available at https://vk.com/doc6752525_146744970?hash=fc2de899bd0582c6b3&dl=a0dc15b757a75af4fe, last accessed on May 18, 2016.

Mari kalyk muro 1951 = V. Koukal' (comp.) *Marii kalyk muro. Mariiskie narodnye pesni.* [Mari Folk Songs.] Leningrad & Moscow: Gosudarstvennoe muzykal'noe izdatel'stvo.

Pesni gornyh mari 2005 = *Pesni gornyh mari. Kyryk mary halyk myryvl*ä. [Songs of the Hill Maris.] Yoshkar-Ola: Mariiskii nauchno-issledovatel'skii institut iazyka, literatury i istorii im. V. M. Vasil'eva.

Podelkov, Sergei 1976. *Mariiskie narodnye pesni.* [Mari Folk Songs.] 2nd ed. Yoshkar-Ola: Mariiskoe knizhnoe izdatel'stvo.

Potebnia, Aleksandr 1989. *Slovo i mif.* [Word and Myth.] Moscow: Pravda.

Ryabina, Elena 2011. Differences in the Distribution of Colour Terms in Colour Space in the Russian, Udmurt and Komi Languages. *ESUKA-JEFUL*, Vol. 2, No. 2, pp. 191–213. Available at http://jeful.ut.ee/public/files/Ryabina%20191-213.pdf, last accessed on May 18, 2016.

Sokolov, Yurii 1941. Russkii fol'klor. [Russian Folklore.] Moscow: Gosudarstvennoye uchebno-pedagogicheskoye izdatel'stvo Narkomprosa RSFSR.

Spengler 2000 = Shpengler, Osval'd. *Zakat Evropy*. [The Decline of the West.] Minsk: Harvest; Moscow: ACT.

Turner, Edith 1997. Anthropology, Symbolic. In: Charlie T. McCormick & Kim Kennedy White (eds.) *Folklore: Encyclopedia of Beliefs, Customs, Tales, Music, and Art.* Volume 1. Santa Barbara, CA & Denver, CO & Oxford: ABC CLIO, pp. 25–29.

Uliashev, Oleg 1999. Tsvet v predstavleniiakh i fol'klore Komi. [Colour in Conceptions and Komi Folklore.] Syktyvkar: Izdatel'stvo URO RAN.

Uusküla, Mari 2007. The Basic Colour Terms of Finnish. *SKY Journal of Linguistics*, Vol. 20, pp. 367–397. Available at http://www.linguistics.fi/julkaisut/SKY2007/UUSKULAo.pdf, last accessed on May 18, 2016.

Uusküla, Mari & Sutrop, Urmas 2007. Preliminary Study of Basic Colour Terms in Modern Hungarian. *Linguistica Uralica*, Vol. 43, No. 2, pp. 102–123. Available at http://www.kirj.ee/11171/, last accessed on May 18, 2016.

Vasilyev 1991 = Vasil'ev, V. M. *Mariiskie narodnye pesni.* [Mari Folk Songs.] Yoshkar-Ola: Mariiskoye knizhnoye izdatel'stvo.

Veselovskii, Aleksandr 1940. *Istoricheskaia poetika.* [Historical Poetics.] Leningrad: Gosudarstvennoe izdatel'stvo Khudozhestvennaia literatura. Available at https://vk.com/doc2967709_437463217?hash=216c150d24a9ca157c&dl=b2ab61387d7bc821bf, last accessed on May 18, 2016.

Vezhbitskaia, Anna. 1996. Oboznacheniya tsveta i universalii zritel'nogo vospriyatiya. [Color Designation and Universals of Visual Perception.] In: M. A. Krongauz (ed.) *Iazyk. Kul'tura. Poznanie*. [Language. Culture. Cognition.] Trans. from English. Moscow: Russkie slovari, pp. 231–291. Available at https://vk.com/doc168973255_426866629?hash=9efe5f9d4ee12791ef&dl=347fb1c2ce232314cf, last accessed on May 18, 2016.

Yanshin 2006 = Ian'shin, Petr. *Psikhosemantika tsveta.* [Psychosemantics of Colour.] Sankt Peterburg: Rech'.

http://dx.doi.org/10.7592/FEJF2016.64.polarity

IDENTIFYING POLARITY IN DIFFERENT TEXT TYPES

Hille Pajupuu, Rene Altrov, Jaan Pajupuu

Abstract: While Sentiment Analysis aims to identify the writer's attitude toward individuals, events or topics, our aim is to predict the possible effect of a written text on the reader. For this purpose, we created an automatic identifier of the polarity of Estonian texts, which is independent of domain and of text type. Depending on the approach chosen – lexicon-based or machine learning – the identifier uses either a lexicon of words with a positive or negative connotation, or a text corpus where orthographic paragraphs have been annotated as positive, negative, neutral or mixed. Both approaches worked well, resulting in a nearly 75% accuracy on average. It was found that in some cases the results depend on the text type, notably, with sports texts the lexicon-based approach yielded a maximum accuracy of 80.3%, while over 88% was gained for opinion stories approached by machine learning.

Keywords: lexicon-based approach, machine learning approach, Naïve Bayes, polarity, sentiment analysis, SVM, text types

INTRODUCTION

The exponential increase in Internet usage and the diverse opportunities to express opinions on the Web have surrounded us with an unprecedented amount of information and different types of texts. To help us find a way through this abundance the recent decade has brought a boom in developing automatic text processing systems, which are able to extract, process and present relevant knowledge (Lloret et al. 2012). One of these systems is Sentiment Analysis (SA)[1], whose main goal is to determine whether a text, or a part of it, is subjective or not and, if subjective, whether it expresses a positive or negative view (Taboada 2016).

SA has mostly been performed for product reviews (incl. movie reviews, hotel reviews), forums, blogs, micro-blogs ("tweets"), news articles and social media in order to disclose the writer's opinions, attitudes and emotions toward individuals, events or topics (Pang & Lee 2008; Ravi & Ravi 2015).

SA has two main approaches to choose from: a lexicon-based approach and a machine learning approach.

The lexicon-based approach assumes that the text contains words with an emotional connotation, which are indicative of the writer's attitude (e.g., *abivalmis* 'helpful' is positive; *ebameeldiv* 'unpleasant' is negative). Analysis of the occurrence of such words in a text shows whether the text's polarity is rather positive or negative. Thus the lexicon-based approach requires a dictionary of words with emotional connotation, where the words carry a negative or positive label. The dictionary may also include information about connotation strength, that is, how positive or negative the word is. The target text is searched for dictionary words, which are respectively annotated in the text. A count of words with either polarity will lead to an assessment of the sentiment orientation of the text. In doing the counting, one may have to consider the possibility that the prior polarity of the word has been changed by context (Taboada 2016; Tabaoda et al. 2011; Wilson et al. 2009).

In the machine learning approach, the polarity of a text is determined by using classifiers. Classifiers are trained on a prepared corpus where documents or sentences have been annotated as either positive and negative, or as positive, negative and neutral. The most used classifiers in SA are Naïve Bayes and the Support Vector Machine (SVM). Classification has been attempted using different text features, such as considering unigrams only, bigrams, a combination of both, incorporating part-of-speech and position information, taking only adjectives, etc. Of all features, unigrams (individual words or tokens) have been the most effective. Accuracy has been increased by including all parts of speech, instead of being confined to adjectives, as well as by considering the position of the word in the text. SVM has generally given better results than Naïve Bayes, but Naïve Bayes also performs excellently, in particular if the feature space is small (Pang et al. 2002; Ravi & Ravi 2015; Taboada 2016; Vegda et al. 2014).

Both the lexicon-based and the machine-learning approach have yielded good results: for English texts, the accuracy is mostly between 70-90%, while higher scores have been achieved with texts sharing a domain (e.g., movie reviews), when subjected to a two-way classification (into positive and negative) (Pang & Lee 2008; Taboada et al. 2011; Zhang et al. 2014).

Which of the two approaches is preferable depends on the availability of sentiment resource, including annotated lexicons and corpora. The main problem with the lexicon-based approach lies in cross-domain adaptability. Words carrying a positive connotation in one domain may be negative or neutral in another. Thus, the lexicon has to be adapted to this or that domain. The machine-learning approach requires much human effort in document annotation and a good match between the training and testing data with respect to the domain (Taboada 2016; Zhang et al. 2014).

The resources are mainly available for English. Few other languages (e.g., Arabic, Spanish, Japanese, Chinese, French) have the necessary corpora or dictionaries. To bridge the gap, there have been attempts to translate some open source English resources into other languages by using the machine translation service. But every language has its peculiarities, which may render the adapting technique unsuitable, so one of the challenges of SA is to create specific linguistic resources for different languages (Balahur et al. 2014; Montoyo et al. 2012; Ravi & Ravi 2015; Taboada 2016; Zhang et al. 2014).

The approaches developed for SA can also be applied to other interlocking tasks.

While SA attempts to pinpoint the writer's attitude toward individuals, events or topics, our objective is to predict the possible effect of a written text on the reader. Unlike SA, where the main task is to tell facts apart from the writer's subjective opinion and analyse the latter, our underlying assumption is that any text, including an objective factual news text will affect the reader's mood, attitudes and decisions. Knowing the objective positive or negative polarity of the text would make this effect predictable.

The need to analyse text polarity ensues from the recent tendency to increasingly replace face-to-face interaction with the written one, both in private and working life. Unlike oral interaction, where the partner's mood and attitude can be detected from their voice and facial expression, written interaction leaves us face to face with nothing but a written text. Problems will arise if the text is misinterpreted. Whenever a writer is not sure what effect their text is likely to produce, an automatic identifier of text polarity could be of help in adjusting the text as necessary.

In addition to interaction, we are daily faced with a huge amount of texts, which makes deciding which ones to read difficult. Any means of automatic text processing to help us make the decision is welcome. One of the criteria motivating our decision is whether the text is positive or negative.

Another motive behind our wish to develop an automatic identifier of the polarity of a written text is a speech-technological need to make the result of text-to-speech synthesis more natural. The selection of the appropriate acoustic model for a text to be voiced requires information on the affectivity of the text (see Tamuri & Mihkla 2015).

Our challenge was to create an automatic identifier of text polarity for the Estonian language.

Estonian belongs to the Finnic branch of the Uralic language family. The language has a rich morphology, thus rendering it rather different from English, which is hitherto the dominating language in polarity identification. Morphological issues may, however, require special attention in polarity identification. For

Arabic, for example, which is also a morphologically rich language, the system includes automatic lemmatization (that is, head form retrieval) and part-of-speech (POS) tagging. This has yielded good results in subjectivity analysis (Abdul-Mageed et al. 2014).

For Estonian, the pilot study attempting to determine the polarity of a written text on the lexicon-based method was done manually on very limited material (Pajupuu et al. 2012a, 2012b). Inspired by the outcome we have set the task to create an automatic identifier of Estonian text polarity, which would be independent of domain and text type, and to find out whether the lexicon-based or machine learning method is more appropriate for the task.

RESOURCES NECESSARY FOR POLARITY IDENTIFICATION

First, a corpus had to be created to train and test the classifiers based on machine learning and to evaluate the accuracy of the lexicon-based method. According to Pajupuu and her colleagues (Pajupuu et al. 2012a, 2012b) it could be assumed that the optimal unit of polarity identification is an orthographic paragraph, not a full document or sentence. A paragraph is mostly a meaningful unit of text, mainly consisting of sentences of a similar polarity (see Figure 1).

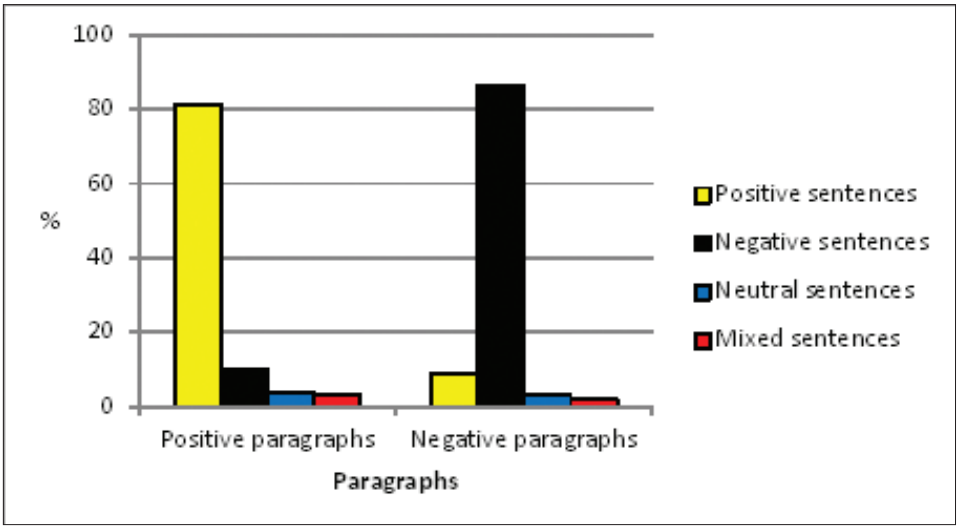

Figure 1. *According to the material of the Estonian Emotional Speech Corpus an orthographic paragraph usually consists of sentences of a similar polarity (Pajupuu et al. 2012b).*

The polarity corpus was compiled of articles of different rubrics of online dailies, weeklies, and reader comments, while the polarity of each paragraph was determined by native Estonian readers. Three subjects were asked to read the paragraphs independently of one another and decide from feeling whether the paragraph is positive, negative, neutral or ambivalent. The paragraphs were annotated using the dominant opinion (Pennebaker et al. 1997). The dominant opinion was the one expressed by at least two of the three readers. If no opinion dominated (all three were different, thus including, for example, positive, neutral and ambivalent), the paragraph was annotated as mixed (see Examples 1–4).

Example 1. A paragraph of an opinion story, annotated in corpus as positive:

Koht, mis varem ei olnud püha, võib selleks saada. Kui istutame tammikud, muudame need kohad pühaks. Hoolitseme ka selle eest, et tammikutes kasvaks kaunis kask ja püha pihlakas, et kaugete esivanemate vaimud end seal hästi tunneksid.[2]
A place that previously was not holy can become like that. We can make it holy ourselves by planting an oak forest. Moreover, let us take care that the oak forest also features the beautiful birch and the protective rowan, just to make the distant ancestral spirits feel good.

Example 2. A paragraph of a crime news story, annotated in corpus as negative:

Tabati ka üks kriminaalses joobes sõidukijuht. See juhtus pühapäeva öösel kella 4 ajal, kui Viljandis Lääne tänaval peeti kinni sõiduauto BMW, mille roolis oli 21-aastane noormees. Tema suhtes alustati kriminaalmenetlust.[3]
Also, a criminally intoxicated driver was apprehended. It happened at 4 o'clock Sunday morning that a BMW driven by a 21-year old was stopped on Lääne St. in Viljandi. Criminal charges were filed.

Example 3. A paragraph of culture news, annotated in corpus as neutral:

Peaaegu samasugune nägi pööning välja märtsis, kui kunstnik oli sinna üles seadnud "Asjade" esimese osa. Vahepealse kuue kuu jooksul on katusealune ja seda külastanud vaatajad osa saanud suurtest muudatustest.[4]
The attic looked almost the same in March, just after the artist had set up the first part of the "Things". During the six months passed, the attic and its visitors have been exposed to some considerable changes.

Example 4. A paragraph of domestic news, annotated in corpus as mixed:

Uuringust tuli välja, et ligi pooled inimesed ei kavatse enam Eestisse tagasi tulla, kuid paljud vastajad tunnistasid, et kui Eestis oleks neil rohkem väljakutseid ja huvitav töö, siis kaaluksid nad tagasitulemist.[5] According to the results, nearly half of the people had no intention of returning to Estonia, however, many respondents admitted that if Estonia offered them more challenges and an interesting job, they would reconsider.

In total, the corpus contains 4,086 annotated paragraphs, see Figure 2 and Table 1. The corpus is an open source and available for free.[6]

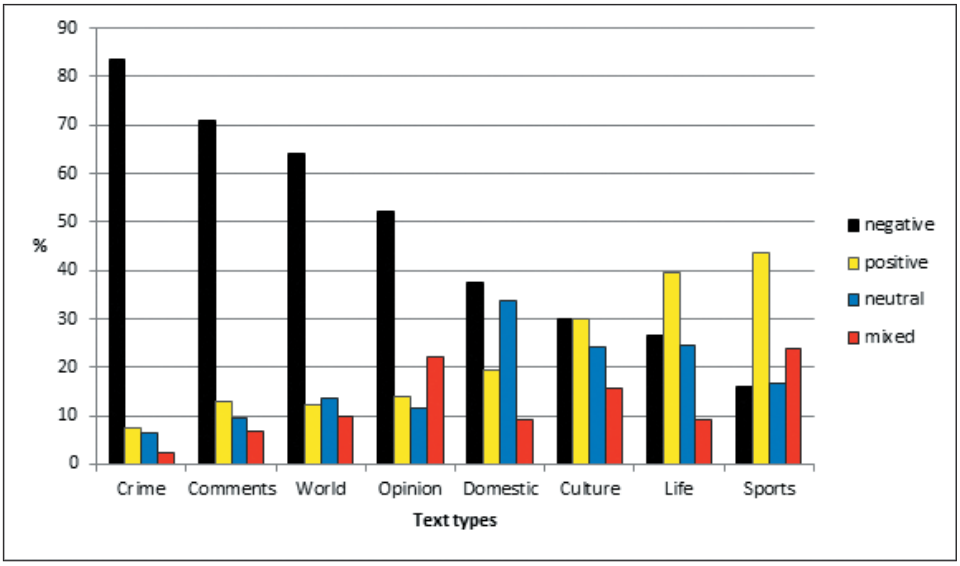

Figure 2. *The annotated corpus paragraphs by text types.*

The lexicon to be used in polarity identification was compiled of words with a positive or negative connotation. We dropped the idea of translating relevant dictionaries from other languages, because every culture has specific words carrying a positive or negative connotation precisely for the members of this particular culture. For a local Estonian person, for example, the positive words

include *leib* 'bread', *vaikne* 'quiet', *sõltumatu* 'independent', whereas *hilinema* 'be late' and *vihmane* 'rainy' are negative. In some other culture the same words may be neutral or even of an opposite connotation.

The volume of the dictionaries used in lexicon-based SA can be very different ranging from the 5,000 polarity-annotated words as in SO-CAL (Taboada et al. 2011) to the nearly 76,000 words of the Macquarie Semantic Orientation Lexicon (Mohammad et al. 2009). The optimal size is still open to discussion. Taboada and her colleagues have found that a large dictionary tends to capture more noise, leading to inaccurate results in SA (Taboada 2016; Taboada et al. 2011). Pajupuu and her colleagues (Pajupuu et al. 2012b) have observed that a relatively small dictionary of frequent words can turn out to be efficient, because most of the frequent polar words are monovalent, so that their connotation is seldom changed by context (e.g., the frequent Estonian word *koostöö* 'cooperation' is invariably positive, whereas the relatively rare word *vähenõudlik* 'undemanding' can be either positive or negative depending on the context).

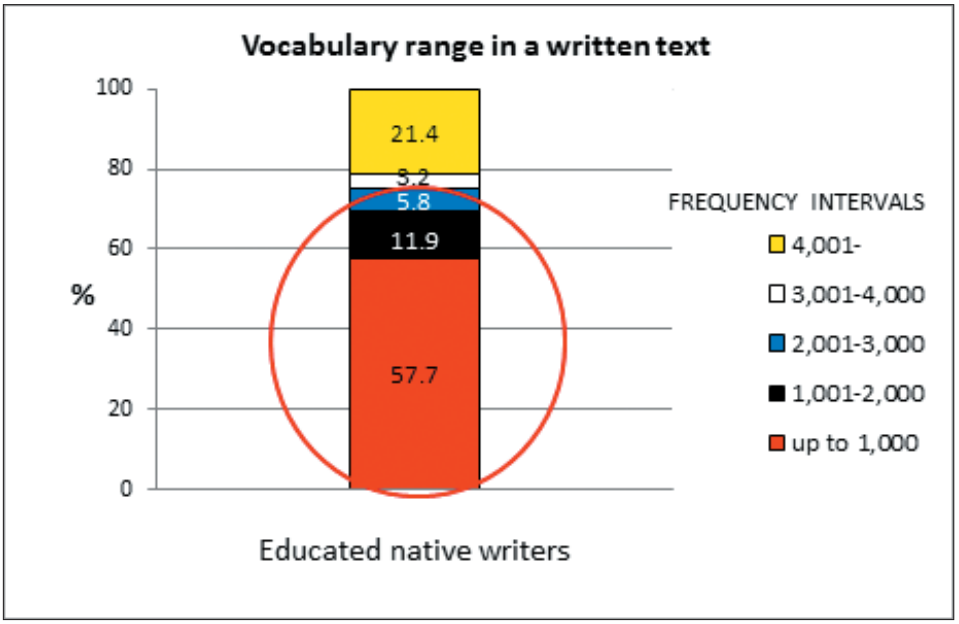

Figure 3. *Vocabulary range in demanding texts by educated native writers (Kerge et al. 2014).*

As our aim was to identify the polarity of texts regardless of their domain and text type, we decided to limit the dictionary to frequent words for the reason that frequent words tend to appear in most texts, whereas rare words rather belong to specific topics or domains. In demanding texts produced by educated writers most (75%) of the words belong to the 3,000 more frequent ones (Kerge et al. 2014), see Figure 3. Hence, if a text has an emotional meaning, it is likely to contain some frequent words with a polarity.

Another question deriving from SA practice is whether the dictionary should only consist of adjectives or should other parts of speech be included (see Taboada 2016). We decided to determine the positive or negative connotation, if any, for all words included in the Basic Estonian Dictionary regardless of their part of speech (see Vainik 2012). The Basic Estonian Dictionary contains the 3,015 most frequent Estonian words. Like in the case of orthographic paragraphs we asked four native speakers of Estonian to determine whether the words have a positive or negative connotation, are neutral or ambivalent (in the latter case polarity depends on context). Due to the dominating opinion, words with a positive (317) or negative (322) connotation were included in the polarity dictionary (cf. Pajupuu et al. 2012b). After supplementing the list with some antonyms and derivatives the polarity dictionary now features 617 words with a positive and 730 words with a negative annotation.

As Estonian is a language of considerable morphological richness and text words can manifest very different grammatical forms, we first developed an idea to use a lemmatizer to reduce all different forms of a word to the dictionary head-word (cf. Abdul-Mageed et al. 2014). The idea was soon dropped, however, because different forms of a word can be of different polarities (e.g., *abi* (pos) 'help' *abita* (neg) 'without help'; *nautima* (pos) 'enjoy', *nautimata* (neg) 'without enjoying'). So we decided to include the words with all their grammatical forms and annotate them as positive or negative. The final size of the dictionary was 38,628 tokens. The addition of grammatical forms revealed cases of morphological homonymy (accidental coincidence between certain grammatical forms of words of different parts of speech) (e.g., *tänavat* (pos) '(he) is known to say thank you' and *tänavat* (neutr) 'this street PartSg'; *lood* (pos) 'you'll create (sth)', *lood* (neutr) 'stories'; *mees* (pos) 'in honey' *mees* (neutr) 'a man'; *mürgita* (pos) 'without poison' and *mürgita* (neg) 'just poison (him)'. Such homonymous forms were excluded from the dictionary.

Next, we used the paragraph corpus to see how many paragraphs contain words included in our dictionary. As the paragraph corpus contains not only positive, negative and mixed paragraphs, but also neutral ones, the results can be regarded as promising, see Table 1.

Table 1. *Paragraph corpus and the number of paragraphs containing at least one word included in the polarity dictionary.*

Text types	**Number of paragraphs**	**Number of paragraphs with a word in the polarity dictionary**	
Opinion	972	876	90.1%
Domestic	419	287	68.5%
Life	518	381	76.3%
Comments	1008	781	77.5%
Crime	209	162	77.5%
Culture	262	183	69.8%
Sports	385	311	80.8%
World	313	239	76.4%

METHOD

In the lexicon-based approach, each text word was compared with those in the polarity dictionary. The following rules were applied:
a) A negation (*ei* or *ega*) will make the following positive word negative, for example, *rõõmustama* (pos), 'to please; rejoice', *ei rõõmusta* (neg) 'will not be/make happy';
b) A negation will make the following negative word positive, for example, *valetama* (neg) 'to lie', *ei valeta* (pos) 'does not lie';
c) A negation will make the following neutral word negative, for example, *tulema* (neutr) 'to come', *ei tule* (neg) 'is not coming'.

Then the positive and negative words were summarized over each paragraph. The paragraph was classified as positive or negative according to the connotation of the dominating words. For example, if a paragraph contained just one positive word (+1) and four negative ones (-4), the paragraph was classified as negative. If the number of positive and negative words was equal, the paragraph was classified as mixed. If there were neither any positive or negative words, the paragraph was classified as neutral (cf. Pajupuu et al. 2012b).

The lexicon-based method was tested on the whole material of the paragraph corpus.

In machine learning the words (unigrams) of the paragraphs were used as features. Two classifiers, Naïve Bayes and the Support Vector Machine (SVM) were tested in the Scikit-learn environment (Pedregosa et al. 2011).

Naïve Bayes was trained using:

```
Pipeline([
    ('vect',CountVectorizer(binary=True)),
    ('clf', MultinomialNB())
    ])
```

SVM was trained using:

```
    Pipeline([
        ('tfidf', TfidfVectorizer(use_idf=True)),
        ('clf', SGDClassifier(loss='squared_hinge', penalty='l2', random_state=None,
        alpha=1e-3))
    ])
```

Both Naïve Bayes and the SVM were trained all over the corpus without discriminating between text types, while a hundred paragraphs of each text type were used for testing. The paragraphs annotated as mixed in corpus were removed from the training material, because in pilot tests such paragraphs were found to have a considerable lowering effect on the accuracy of identification. A three-way classification (positive, negative, neutral) was used.

In order to evaluate the polarity classification accuracy of the lexicon-based method versus machine learning their outputs were compared with the human ratings of the corpus paragraphs. Automatic classification was considered correct if it coincided with the human one, plus the positive–mixed and negative–mixed pairs.[7]

From the polarity of the orthographic paragraphs found either by the lexicon-based or machine learning method, the polarity of the full document was determined. First, the words in a paragraph were counted. Each word was assigned the polarity of the paragraph it belonged to, that is, if a paragraph was positive (negative, neutral, mixed), all its words got a respective positive (negative, neutral, mixed) label.

If a document consisted of paragraphs of one and the same class, the document was assigned the same polarity[8]: POSITIVE, NEGATIVE, NEUTRAL or MIXED.

If a document contained paragraphs of two classes, the proportion of the words of either class in the document was calculated. If the words of a class (positive, negative, neutral, or mixed) made up at least 66.7% of the document, the latter was respectively labelled as MOSTLY POSITIVE, MOSTLY NEGATIVE, MOSTLY NEUTRAL or MOSTLY MIXED. If the words of none of the classes reached 66.7% of the document words, the document was labelled as MOSTLY MIXED.

If a document contained paragraphs of three or four classes, the document was labelled after the class whose words made up at least half of the document words (MOSTLY POSITIVE, MOSTLY NEGATIVE, MOSTLY NEUTRAL or MOSTLY MIXED). In the rest of cases the document was labelled as MOSTLY MIXED.

The accuracy of the labelling of full documents was not evaluated, because the corpus does not include full documents with human-determined or human-annotated polarity.

RESULTS AND DISCUSSION

Our aim was to create an automatic identifier of Estonian text polarity, which would be independent of domain and text type, and find out whether the lexicon-based or machine learning approach should be preferred in doing so.

Table 2 presents, by text types, the percentage of paragraphs with accurately identified polarity as found by using the lexicon-based approach versus the machine learning method using the classifiers Naïve Bayes and SVM.

Table 2. *Percentage of paragraphs with accurately identified polarity, by text types.*

Text types	**Number of paragraphs**	**Lexicon-based approach**	**Machine learning approach**			
			Naïve Bayes		**SVM**	
			M	SD	M	SD
Opinion	972	76.4	88.8	1.8	88.2	1.3
Domestic	419	73.0	62.7	1.7	67.2	1.8
Life	518	62.4	64.5	2.7	69.5	2.4
Comments	1008	58.3	86.0	2.9	85.2	1.4
Crime	209	78.0	87.9	2.0	87.7	1.2
Culture	262	78.2	54.4	5.2	58.1	2.3
Sports	385	80.3	75.7	3.0	75.2	2.6
World	313	79.2	68.7	1.1	72.5	2.5
M		*73.2*	*73.6*		*75.5*	
SD		*8.3*	*13.0*		*10.8*	

Note. *For Naïve Bayes and SVM an average of five tests is presented.*

In the lexicon-based approach the paragraphs were classified into four polarity classes (positive, negative, neutral, mixed). The accuracy was between 58.3–80.3, being lowest for comments and highest for sports articles. In the machine

learning the paragraphs were classified into three polarity classes (positive, negative, neutral). Naïve Bayes yielded accuracy readings from 54.4–88.8, the lowest score belonging to cultural texts and the highest to opinion stories. With SVM the accuracy ranged from 58.1–88.2, being also lowest for cultural texts, but the highest accuracy was scored by opinion stories. There was no substantial difference between the two approaches in polarity identification (the average score being 73.2–75.5), but the accuracy scores did differ across text types.

In the lexicon-based approach, the lowest accuracy was measured in the case of comments. One of the reasons is probably that comments are often very short, consisting of a single sentence or just a word, whereas the optimal unit for lexicon-based identification of polarity is an orthographic paragraph, that is, at least two semantically connected sentences. A single sentence, however, may not contain any words from the polarity dictionary and so the sentence will be classified as neutral. For example, the sentence *Poodi on tarvis, aga ehitage kellegi teise kodu kõrvale* (A store is necessary, but build it next to someone else's home) was classified as negative by humans, but neutral by the lexicon-based identifier. Moreover, comments often feature unusual word usage (e.g., swear words, abbreviations, slang, foreign words and expressions) and deviation from regular orthography. If a word is relatively rare and has deviant orthography, it cannot be found in our polarity dictionary and will consequently be regarded as neutral (e.g., the two-word sentence *krdi ajukääbik* (you damned fool), which consists of an abbreviation and a rare derogatory word, was classified as negative by humans, but neutral by the lexicon-based identifier). The machine learning method, however, copes well with comments, gaining an accuracy of 86.0% with Naïve Bayes and 85.2% with SVM.

In machine learning the polarity identification accuracy was the lowest with paragraphs about culture. Further analysis should disclose whether the reason could lie in the small number of culture paragraphs in the training corpus, and in their lexical diversity. Actually, crime, with a still smaller number of paragraphs and low lexical variation, ended up with a very high accuracy indeed.

Although our aim was to find out which approach, the lexicon-based one or machine learning, works better for the identification of polarity in Estonian texts, the answer is still ambiguous. According to the mean accuracy (~75%) both approaches can be regarded as equally appropriate and worthy of further development. The lexicon-based approach requires the existence of a polarity dictionary, while machine learning requires a corpus of polarity-annotated texts (see e.g., Balahur et al. 2014; Taboada 2016). Both are now available for the Estonian language, open for public use, improvement and extension.

The results of polarity identification for Estonian written texts are not quite comparable with the SA results available for some other languages. SA has a different purpose, notably, to identify the writer's attitude towards an entity or topic (see e.g., Montoyo et al. 2012). Our aim was to identify the polarity of a text in order to predict its possible effect on the reader. Any text, however subjective or objective, can carry a positive or negative meaning (cf. Patel et al. 2015). On the one hand, the task of SA is somewhat more complicated in that first, one has to discern and separate the subjective part of the text and then identify the polarity of this part, leaving the objective part aside. On the other hand, however, SA looks rather more simple, because it is mostly domain and text-type centred and in many cases the classification used is dichotomous, dividing its objects into positive and negative ones (see Ravi & Ravi 2015). Our identifier is neither domain or text-type centred and the objects are divided either between four classes (as in the lexicon-based approach) or between three classes (as in machine learning). Thus, for us the 70–90% accuracy rates of SA prevalently scored on English material can be regarded as an approximate benchmark only. Our mean accuracy of ~75% is two to three times better than chance probability, which can be considered a sufficiently good score for a polarity identifier that is independent both of text type and of domain.

Little is known of SA studies for languages with a rich morphology. For SA performed on Arabic social media texts a lemmatizer and POS-tagging were used and their dichotomous classification into positive and negative yielded accuracies of 70.3–81.8%, depending on the text type (Abdul-Mageed et al. 2014). Our polarity identifier, using a three-way and a four-way classification, gave a mean accuracy of the same interval. Whether lemmatization and POS-tagging could raise the accuracy even more remains to be tested. Our accuracy currently achieved in polarity identification by the machine learning method is consistent with the results used to prove that the simple use of unigrams as features leads to good results (cf. Balahur et al. 2014; Pang & Lee 2008; Vegda et al. 2014).

Our results from the lexicon-based approach corroborate the statement of Taboada et al. (2011) that small dictionaries can do very well for SA. Their nearly 5,000-word dictionary worked better in their Semantic Orientation CALculator (SO-CAL) (mean accuracy 78.7%) than bigger dictionaries. Our polarity identifier, using our polarity dictionary of 1,347 frequent words with a positive or negative connotation performed similarly well. As far as we know, our dictionary is one of the smallest used in this field. As the compilation of a polarity dictionary of frequent words of this amount is a relatively simple task requiring little human resources, whatever the language, it is well worth giving it a try.

Our success is largely due to the introduction of the orthographic paragraph as a unit of identification. If a person labels a paragraph as negative or positive, the paragraph usually contains some frequent words of a negative or positive connotation, respectively, enabling our lexicon-based polarity identifier to perform, in several text domains, rather similarly to humans.

The limitation of the present study is the relatively small number of domains and text types involved. There may still be domains and text types requiring an upgrade of both the training corpus and the dictionary. The necessity is implied by the relatively lower accuracy of comments polarity identification in the case of the lexicon-based approach. The written texts whose style resembles that of oral speech (real time messages, blog, tweets, comments etc.) certainly need the polarity dictionary to be supplemented with frequent polar words characteristic of their text types.

By way of conclusion, we have created resources for polarity identification in the Estonian language (one of the Finnic branch of the Uralic language family) and tested both the lexicon-based and the machine learning approaches to the classification of texts by polarity[9]. The results look promising and the problems revealed are worth further investigation, including in the SA direction.

ACKNOWLEDGEMENTS

This work was supported by institutional research funding IUT 35-1 of the Estonian Ministry of Education and Research, TK 145 "Centre of Excellence in Estonian Studies – CEES", and the governmental basic financing of the Institute of the Estonian Language from the Estonian Ministry of Education and Research.

NOTES

[1] Sentiment Analysis, Opinion Mining, and Subjectivity Analysis are broadly used as synonyms, although some sources make a difference between the concepts (Pang & Lee 2008; Serrano-Guerrero et al. 2015).

[2] Quoted from Sutrop, Urmas. Nagu Taara tammikud. [Like Taara oak-woods.] Newspaper *Maaleht*, December 19, 2014. Available at http://maaleht.delfi.ee/news/maaleht/arvamus/urmas-sutrop-nagu-taara-tammikud?id=70389341, last accessed on May 20, 2016.

[3] Quoted from Teder, Merike. Tabatud joobes juhid saadeti arestimajja. [Caught drunk drivers were sent to detention.] Newspaper *Postimees*, September 11, 2012. Available at http://www.postimees.ee/968740/tabatud-joobes-juhid-saadeti-arestimajja, last accessed on May 19, 2016.

[4] Quoted from Hanson, Raimu. Pööningu kilast-kolast käis üle jumalik hingus. [Divine breath flew over the junk in the attic.] Newspaper *Tartu Postimees*, September 13, 2012. Available at http://tartu.postimees.ee/970946/pooningu-kilast-kolast-kais-ule-jumalik-hingus, last accessed on May 19, 2016.

[5] Quoted from Traks, Kristina. Mis tooks Eesti töötajad Soomest tagasi? [What would bring Estonian workers back from Finland?] Newspaper *Postimees*, September 13, 2012. Available at http://majandus24.postimees.ee/971422/mis-tooks-eesti-tootajad-soomest-tagasi, last accessed on May 19, 2016.

[6] Free and open corpus of paragraphs http://peeter.eki.ee:5000/valence/paragraphsquery/.

[7] An orthographic corpus paragraph has been annotated as mixed in two cases: (1) the readers have determined the paragraph as ambivalent, which means that it contains both positive and negative elements; (2) there is no dominant opinion (e.g., half of the readers have determined the paragraph as negative, the other half as positive). Therefore, we decided that a "mixed" paragraph can be considered correctly identified by the program if it classifies the paragraph as positive or negative.

[8] Polarity in a wider sense includes four classes: positive, negative, neutral and mixed.

[9] Automatic identifier of written text polarity, http://peeter.eki.ee:5000/valence/, and https://github.com/EKT1/valence/.

REFERENCES

Abdul-Mageed, Muhammad & Diab, Mona & Kübler, Sandra 2014. SAMAR: Subjectivity and Sentiment Analysis for Arabic Social Media. *Computer Speech and Language*, Vol. 28, No. 1, pp. 20–37. http://dx.doi.org/10.1016/j.csl.2013.03.001.

Balahur, Alexandra & Mihalcea, Rada & Montoyo, Andres 2014. Computational Approaches to Subjectivity and Sentiment Analysis: Present and Envisaged Methods and Applications. *Computer Speech and Language*, Vol. 28, No. 1, pp. 1–6. http://dx.doi.org/10.1016/j.csl.2013.09.003.

Kerge, Krista & Pajupuu, Hille & Alp, Pilvi & Põlda, Halliki & Uusen, Anne 2014. Towards Sophisticated Writing. *Proceedings of the Tallinn University Institute of Estonian Language and Culture: Studies in Language Acquisition, Learning, and Corpora*, Vol. 16, pp. 103–115. Available at http://www.academia.edu/9729151/Towards_sophisticated_writing, last accessed on May 19, 2016.

Lloret, Elena & Balahur, Alexandra & Gómez, José & Montoyo, Andrés & Palomar, Manuel 2012. Towards a Unified Framework for Opinion Retrieval, Mining and Summarization. *Journal of Intelligent Information Systems*, Vol. 39, No. 3, pp. 711–747. http://dx.doi.org/10.1007/s10844-012-0209-4.

Mohammad, Saif & Dorr, Bonnie & Dunne, Cody 2009. Generating High-Coverage Semantic Orientation Lexicons from Overtly Marked Words and a Thesaurus. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*, August 6–7, Singapore, pp. 599–608.

Montoyo, Andrés & Martínez-Barco, Patricio & Balahur, Alexandra 2012. Subjectivity and Sentiment Analysis: An Overview of the Current State of the Area and Envisaged Developments. *Decision Support Systems*, Vol. 53, No. 4, pp. 675–679. http://dx.doi.org/10.1016/j.dss.2012.05.022.

Pajupuu, Hille & Kerge, Krista & Altrov, Rene 2012a. Detecting Emotional Valence of Text by Using a Small Dictionary. In: Izaskun Elorza & Ovidi Carbonell i Cortés & Reyes Albarrán & Blanca García Riaza & Miriam Pérez-Veneros (eds.) *Empiricism and Analytical Tools for 21st Century Applied Linguistics. Selected Papers from the XXIX International Conference of the Spanish Association of Applied Linguistics (AESLA)*, Collección Aquilafuente 185. Salamanca: Universidad de Salamanca, pp. 229–242.

Pajupuu, Hille & Kerge, Krista & Altrov, Rene 2012b. Lexicon-Based Detection of Emotion in Different Types of Texts: Preliminary Remarks. *Eesti Rakenduslingvistika Ühingu aastaraamat / Estonian Papers in Applied Linguistics*, Vol. 8, pp. 171–184. DOI:10.5128/ERYa8.11.

Pang, Bo & Lee, Lillian 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, Vol. 2, Nos. 1–2, pp. 1–135. http://dx.doi.org/10.1561/1500000011.

Pang, Bo & Lee, Lillian & Vaithyanathan, Shivakumar 2002. Thumbs Up?: Sentiment Classification Using Machine Learning Techniques. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Vol. 10, pp. 79–86. DOI:10.3115/1118693.1118704.

Patel, Vishakha & Prabhu, Gayatri & Bhowmick, Kiran 2015. A Survey of Opinion Mining and Sentiment Analysis. *International Journal of Computer Applications*, Vol. 131, No. 1, pp. 24–27. http://dx.doi.org/10.5120/ijca2015907218.

Pedregosa, Fabian & Varoquaux, Gaël & Gramfort, Alexandre & Michel, Vincent & Thirion, Bertrand & Grisel, Olivier & Blondel, Mathieu & Prettenhofer, Peter & Weiss, Ron & Dubourg, Vincent & Vanderplas, Jake & Passos, Alexandre & Cournapeau, David & Brucher, Matthieu & Perrot, Matthieu & Duchesnay, Édouard 2011. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830. Available at http://www.jmlr.org/papers/v12/pedregosa11a.html, last accessed on May 19, 2016.

Pennebaker, James W. & Mayne, Tracy J. & Francis, Martha E. 1997. Linguistic Predictors of Adaptive Bereavement. *Journal of Personality and Social Psychology*, Vol. 72, No. 4, pp. 863–871. http://dx.doi.org/10.1037/0022-3514.72.4.863.

Ravi, Kumar & Ravi, Vadlamani 2015. A Survey on Opinion Mining and Sentiment Analysis: Tasks, Approaches and Applications. *Knowledge-Based Systems*, Vol. 89, pp. 14–46. http://dx.doi.org/10.1016/j.knosys.2015.06.015.

Serrano-Guerrero, Jesus & Olivas, Jose A. & Romero, Francisco P. & Herrera-Viedma, Enrique 2015. Sentiment Analysis: A Review and Comparative Analysis of Web Services. *Information Sciences*, Vol. 311, pp. 18–38. DOI:10.1016/j.ins.2015.03.040.

Taboada, Maite 2016. Sentiment Analysis: An Overview from Linguistics. *Annual Review of Linguistics*, Vol. 2, pp. 325–347. http://dx.doi.org/10.1146/annurev-linguistics-011415-040518.

Taboada, Maite & Brooke, Julian & Tofiloski, Milan & Voll, Kimberly & Stede, Manfred 2011. Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, Vol. 37, No. 2, pp. 267–307. http://dx.doi.org/10.1162/COLI_a_00049.

Tamuri, Kairi & Mihkla, Meelis 2015. Expression of Basic Emotions in Estonian Parametric Text-To-Speech Synthesis. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics*, Vol. 6, No. 3, pp. 145–168. http://dx.doi.org/10.12697/jeful.2015.6.3.06.

Vainik, Ene 2012. Kuidas määrata eesti keele sõnavara tundetoone? [Detecting Emotional Valencies for the Estonian Vocabulary.] *Eesti Rakenduslingvistika Ühingu aastaraamat / Estonian Papers in Applied Linguistics*, Vol. 8, pp. 257–274. DOI: http://dx.doi.org/10.5128/ERYa8.17.

Vegda, Hiteshkumar N. & Patel, Tejal R. & Patel, Bhargesh B. 2014. A Survey on Sentiment Analysis of Texual (sic) Review. *International Journal of Research in Advent Technology*, Vol. 2, No. 2, pp. 2321–9637. Available at http://www.academia.edu/9310690/A_survey_on_Sentiment_Analysis_of_Texual_Review, last accessed on May 19, 2016.

Wilson, Theresa & Wiebe, Janyce & Hoffmann, Paul 2009. Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis. *Computational Linguistics*, Vol. 35, No. 3, pp. 399–433. http://dx.doi.org/10.1162/coli.08-012-R1-06-90.

Zhang, Hailong & Gan, Wenyan & Jiang, Bo 2014. Machine Learning and Lexicon Based Methods for Sentiment Classification: A Survey. *11th Web Information System and Application Conference*, 12–14 September 2014, Tianjin, China, pp. 262–265. DOI: 10.1109/WISA.2014.55.

http://dx.doi.org/10.7592/FEJF2016.64.piits

MAN'S GENDER AND AGE AS BASED ON THE COLLOCATIONS OF THE ESTONIAN WORD *MEES* 'MAN'

Liisi Piits

Abstract: This study explores collocational patterns of the noun *mees* 'man' (in comparison with the data found for *naine* 'woman' and *poiss* 'boy'), which can reveal social attitudes and stereotypes of age, gender and behaviour. The corpus based study uses the Estonian Reference Corpus (203 million words altogether) where 75% of the texts are journalistic, so the results will mainly show how people are depicted in the media. Research data was retrieved by means of the Sketch Engine (see Kilgarriff et al. 2004), which enables collocations to be found using lexico-grammatical constructions. Two lexico-grammatical constructions were analysed: the adjectival modifiers of nouns and verbs co-occurring with a noun functioning as a subject.

Differences in the descriptions of men, women and boys as reflected in specific adjectives and verbs are discussed under five themes: man's aggressiveness, man in verbal interaction and his emotions, man in public and private spheres, man's appearance and personality and man as an adult.

Keywords: age, adjectives, boy, collocation, Estonian, gender, man, verbs, woman

INTRODUCTION

As it is generally understood, the semantic process in mother tongue acquisition cannot happen without contextual (general as well as linguistic) support, i.e. an understanding of a word's meaning develops and is fixed in the process of use. If this is the case, the true meaning of a word could perhaps be discovered by an opposite movement from a big text corpus, over collocations, to the word. It has been argued that collocational patterns can reveal social attitudes, particularly with regard to certain stereotypes of age, gender and behaviour (Caldas-Coulthard & Moon 2010). Thus, we proceed on the assumption that frequent collocational relations enable one to deconstruct latent meanings and analyse how men are (subconsciously) depicted in texts.

The topic of the article is among those discussed in my doctoral dissertation *Collocations of the most frequent Estonian words for 'human being'* (Piits 2015),

supervised by professor Urmas Sutrop. The present study, however, is focused on the adjectival modifiers of the word *mees* 'man' and the verbs for which *mees* acts as a subject (in comparison with the data found for gender-contrasted *naine* 'woman' and for age-contrasted *poiss* 'boy'). Analysis of the adjectival modifiers is expected to demonstrate the typical description of men, while the frequent verb collocates should reveal the activities associated with men as subjects. My main focus is on the collocational relations specific to the word *mees* 'man', concentrating on those adjectives and verbs whose co-occurence with *mees* is at least three times higher than with the gender-contrasted *naine* 'woman' or with the age-contrasted *poiss* 'boy'. This is expected to reveal the specific traits of age and gender that are attributed to men by the users of the Estonian language. However, despite the spotlight being on differences between men and women or men and boys, one should not forget the huge amount of shared collocates. The word *mees* 'man' has a thousand adjectival attributes; 36% of them occur at least once among the attributes of *poiss* 'boy' as well, while the coincidence between the adjectival attributes of *mees* 'man' and *naine* 'woman' is as high as 60%. For my dissertation I also collected material for words like *tüdruk* 'girl', *poeg* 'son', *tütar* 'daughter', *laps* 'child' etc. But in this article I focus on collocations of word *mees* 'man' and collocations of its gender- and age-contrasted antonyms.

The article makes a difference between sex and gender, the former denoting the anatomical and physiological differences between men and women and the latter standing for cultural or learnt differences. Ever since the 1970s when the US anthropologist Gayle Rubin introduced the terminological division, there has been no consensus on the matter. Judith Butler (1990), for example, argues that the notion of biological sex is often culturally determined. According to several scholars, some statements of natural science also have a gendered basis, and, thus, the notion of a binary division of sexes has undergone a historical change (Liljeström 2003: 118). In spite of this, in the present study, a special note will be made if collocates directly depend on the sex of the node word referent, such as, for example, verbs of reproductive functions, like *sünnitama* 'to give birth', *viljastuma* 'be fertilised', *viljastama* 'to fertilise' and *rasestama* 'to impregnate', or adjectives of fertility, like *rase* 'pregnant', *(last) ootav* 'expecting (a baby)' and *imetav* 'breastfeeding'. And yet, we are more interested in gender reflections in language as language is, to a great extent, where gender is born, where gender roles and stereotypes are reinforced and reproduced. Gender is, after all, defined as an assemblage of stereotypes or people's beliefs about a group and its members. Gender stereotypes are manifested in various aspects of life, such as personality traits, behaviour, professional choices, hobbies, looks, family functioning and various preferences (Huntoon 2009: 378).

Apart from gender specifics, the second spotlight of the article is on the possible age specifics of men. According to the Estonian Explanatory Dictionary (EKSS 2009), the major sense of the word *mees* 'man' is *meessoost inimene*, *meesisik*, *meesterahvas* 'male person', from which it could be assumed that *mees* applies indiscriminately to all age groups. However, sometimes the word could be used as opposed to *naine* 'woman' or *poiss (poisike)* 'boy (little boy)'. Opposition to boy as a non-adult should reveal the adult traits and adult activities of men. Although the development from boyhood to adulthood occurs gradually and while its interpretation depends on whether we view it from a religious, socio-cultural or juridical perspective (Panther & Thornburg 2012: 2), some differences can be expected to manifest.

DATA AND TOOLS

The linguistic data were collected from the Estonian Reference Corpus maintained by the University of Tartu. Most of the corpus (203 million words altogether) consists of journalistic texts, but texts of fiction, law, chat rooms, web forums and scientific articles are also represented. The fiction texts are not older than the 1990s, and the journalistic texts represent the period from 1995–2008. Journalistic material is collected from daily newspapers *Postimees, Päevaleht* and Õhtuleht as well as from weekly newspapers *Eesti Ekspress, Maaleht* and different magazines like *Kroonika*, *Horisont* and *Eesti* Arst. The scientific articles come from 1997–2006. The new media texts were downloaded from the Internet from 2003–2008. For a detailed description of the Reference Corpus, see the homepage at http://www.cl.ut.ee/korpused/segakorpus. As the corpus is not a balanced one, as 75% of the texts are journalistic, the results will mainly show how people are depicted in the media. According to Barbi Pilvre (2011a), for example, the media has an important role in the development of stereotypes. She notes that, in the media, women tend to be depicted in certain stereotyped roles associated with the private sphere; also, aspects of the female body are often described in personal portraits of the weekly *Eesti Ekspress*, as well as in other media, when covering women (Pilvre 2011a: 182; Pilvre 2011b). The media in general has been criticised for the advance of a person-centred approach (Nuolijärvi & Tiittula 2000: 19). As for style, the media tends to be dominated by adult, mostly educated usage. It is not impossible that a higher proportion of fiction or new media texts would have produced different results.

Research data was retrieved by means of the Sketch Engine (see Kilgarriff et al. 2004), which enables collocations to be found using lexico-grammatical

constructions. The program recognises collocations on the basis of the output of automatic morphological analysis (part of speech plus grammatical features) combined with statistical methods. Evaluators of the Sketch Engine have noted that the quality of the results depends not only on the size and quality of the corpus and the adequacy of morphological analysis but to a considerable extent also on the word sketch grammar used (Kilgarriff et al. 2010: 4).

The present study uses the word sketch grammar devised by Jelena Kallas (Version 1.5, for a detailed survey of the rules see Kallas 2013). The program searches the data for grammatically linked word pairs and ranks them according to their frequency of co-occurrence. Two lexico-grammatical constructions were analysed:

1. The adjectival modifiers of nouns were detected by the rule called *Adj_modifier* in the Estonian module of the Sketch Engine (see Kallas 2013: 36), which uses part-of-speech labels to pick adjectives of the positive degree of comparison, in whatever case, governed by the following noun in case and number. The use of the adjectival modifiers thus found should be a straight indication of how man is typically described.
2. The verbs co-occurring with a noun functioning as subject were found by the rule *subject_of* (see Kallas 2013: 64), which recognises the finite and personal forms of verbs co-occurring with a full subject. The frequent verbs whose subject is *mees* 'man' show for what activities men are typically described as subjects.

It is true that the rules cannot be expected to recognise all of the relevant cases, and some additional mistakes may occur due to morphological analysis errors and partly inefficient rules (see Piits 2015: 79–80; 103–105), but, as the rules applied are the same for all words, the output adjectives and verbs are comparable enough. Table 1 presents the Corpus frequencies of the words *mees* 'man', *naine* 'woman' and *poiss* 'boy' and the numbers of verbs and adjectival modifiers, together with their percentage of the node word frequency of occurrence. For example, in 14% of the occurrences, the word *mees* 'man' co-occurs with and adjectival modifier while, in 17% of the occurrences, the word *mees* has the function of a subject.

Table 1. *Corpus frequency of the words mees 'man', naine 'woman' and poiss 'boy' and the number of constructions detected by the Adj_modifier and subject_of rules.*

Node word	**Corpus frequency**	**No. of adjectival modifiers and their percentage of node word frequency**	**No. of verb constructions and their percentage of node word frequency**
mees 'man'	307624	43922 (14.0%)	52914 (17.2%)
naine 'woman'	179449	23997 (13.0%)	26160 (14.6%)
poiss 'boy'	59315	10943 (18.0%)	10705 (18.0%)

Of the three words examined, *mees* 'man' has the highest Corpus frequency reading, which is 1.7 times higher than that of *naine* 'woman' and 5.2 times higher than that of *poiss* 'boy'. The difference between the frequency readings of *mees* 'man' and *naine* 'woman' proves the rule that words of male reference are more frequent than those of their female counterparts. A similar tendency has been observed in other languages, such as, for example, English, where the lemma *man* is 1.5 times more frequent than the lemma *woman* (Pearce 2008: 2). The fact that *poiss* 'boy' is several times less frequent than *mees* 'man' obviously shows that non-adults and their activities are not a particularly important subject for journalistic texts. In order to enable the comparability of the collocation frequencies of node words with different frequencies of occurrence, I computed a normalised co-occurrence rate, indicating the ratio of the frequency of a specific collocate to all other collocations found with the node word in question. The use of normalised co-occurrence levels down the differences conditioned by the incidence differences of the node words (see Biber et al. 1998: 263, 264). In order to avoid accidental results, I left aside all collocates occurring less than four times in every 10000 verb constructions or adjective extensions. Moreover, as the aim of the present study was to shed light on differences, my analysis was only applied to those collocates whose normalised co-occurrence rate with the node word is at least three times higher than with another node word. Detailed data about collocation frequencies of the nouns *mees*, *naine* and *poiss* is accessible in http://hdl.handle.net/10062/45486, where the file *Lisa 5* contains the data this article is based on.

AGGRESSIVE MAN

In the description of men, adjectives associated with criminal or deviant behaviour are very salient. Adjectives such as *kahtlane* 'suspicious', *süüdistatav* 'accused', *kahtlustatav* 'suspected', *vägivaldne* 'violent', *verine* 'bloody' and *tundmatu* 'unknown' occur much more often among the collocates of *mees* 'man' than among those of *poiss* 'boy' or *naine* 'woman'. Among the collocates of *naine* 'woman', the word *tundmatu* 'unknown' is found four times more seldom than with *mees* 'man'. Curiously, in more than half of the cases, the phrase *tundmatu mees* 'unknown man' occurs in the descriptions of a perpetrator while 75% of the uses of *tundmatu naine* 'unknown woman' apply to a victim. The adjective *verine* 'bloody' is mostly found in the descriptions of the male victim of a crime or a traffic accident.

Although at first sight the man-specific *rääkiv* 'speaking' and *kõnelev* 'speaking' seem indicative of men's overwhelming need of expression, a closer inspection of the concordance lines revealed that these words belong to the description of perpetrators and victims in the police section. In addition, nearly always, these words occur in longer phrases with modifiers of language competence, such as *eesti* / *vene keelt rääkiv* / *kõnelev* 'Estonian/Russian-speaking' (see examples).

> 1. *Peksjad olid noored* ***eesti keelt rääkivad mehed***.
> The beaters were young Estonian-speaking men. (*Eesti Päevaleht* 13.11.1998)

> 2. *Kaks* ***vene keelt kõnelevat meest*** *on röövinud tänaval relvataolise esemega* ähvardades üksikuid *naisi*.
> Two Russian-speaking men have been robbing single women, threatening them in the street with a gun-like object. (*SL Õhtuleht* 03.10.1999)

In eight cases out of ten, reference is made to Russian speakers while Estonian is mentioned once and so is a third language. Why? The reading cannot be explained just by most of the criminals having a Russian-speaking background (no more but 57% of Estonian prisoners speak Russian as their mother tongue while 40% are native speakers of Estonian (Ahven & Kruusmaa 2013: 104)). The reason should be rather seen in markedness. In Estonia, official Estonian is obviously the norm, not an aberration worthy of verbal marking.

Verb analysis also shows that men are mentioned as subjects of aggressive activities many more times than are women. The verbs *vägistama* 'rape', *tungima* 'intrude', *kaaperdama* 'hijack', *tulistama* 'shoot', *röövima* 'rob', *pooma*

'hang' and *võitlema* 'battle; fight', for example, are specific to men. No aggressive verbs specific to women were revealed.

The description of boys implies a similar degree of aggressiveness, but the activities mentioned are different. Non-adult deviant behaviour includes teasing, bullying, fighting, breaking and hijacking, whereas adult men are mentioned as subjects of *vägistama* 'rape', *tungima* 'intrude', *pussitama* 'stab' and *petma* 'deceive' several times more often than boys.

According to Pearce's analysis of English data (British National Corpus), words associated with crime and offence also collocate more strongly with *man* than with *woman* (Pearce 2008: 9). The asymmetry looks reasonable as most of the perpetrators are male. In Estonia, male persons make up 88% of suspects (Lindsalu 2012: 8) and 95% of prisoners (Ahven & Kruusmaa 2013: 103). Most of the victims are also male.

MAN IN VERBAL INTERACTION AND HIS EMOTIONS

According to Janet Hyde's (2005) gender stereotypic view, women are more emotional and verbally more able than men, although she fails to corroborate her statement with valid empirical proof. In contrast, Hyde presents results to the effect that male/female differences in verbal ability and behaviour are small. We also have no comparative data on the verbal abilities of Estonian men and women, yet a closer look at the 30 most frequent gender-specific collocate verbs of words for 'human being' arranged in antonym pairs, such as *mees-naine* 'man-woman', *isa-ema* 'mother-father', *poeg-tütar* 'son-daughter' and *poiss-tüdruk* 'boy-girl', reveals that the collocates of words with a female referent include more verbs of interaction than do those with a male referent (Piits 2015). The number of gender-specific verbs thus revealed did not differ within the word pairs. The verbs gender-specific to *mees* 'man' are *käsutama* 'command', *vaidlustama* 'contest', ähvardama 'threaten', *vestlema* 'converse', *muhelema* 'chuckle' and *pomisema* 'mumble'. The first three refer to a verbal act for which the subject has an ability as well as wish to use his power while *pomisema* 'mumble' is an interaction verb whose subject does not wish to exert an active influence on others, let alone enforce his power. Although the specific verbs of *naine* 'woman' also include those whose subjects wish to interfere, such as, for example, *manitsema* 'admonish' and *pahandama* 'scold', their subjects lack power. In addition, verbs specific to women include *nuuksuma* 'sob', *ahastama* 'despair' and *karjatama* 'scream', which indicate emotions mainly occurring in victims or persons in a powerless position. Besides the verbs of

emotional interaction, the woman-specific lexis includes a number of other emotionally loaded words, such as, for example, *kiljuma* 'shriek', *ehmuma* 'be startled', *pelgama* 'fear', *vihkama* 'hate' and *ihaldama* 'desire', while the word man collocates only with *muhelema* 'chuckle' (both gender and age specific) and the age-specific *vihastama* 'get angry'.

Age-specific (compared to boys) interaction verbs include *käsutama* 'command', *vaidlustama* 'contend' and *süüdistama* 'blame', which indicate a power position; in addition, there are *kiruma* 'curse' and *vanduma* 'swear' and such adult style activities as *vestlema* 'converse', *nentima* 'state' and *avaldama* 'declare'. Obviously, men converse, state and declare, while boys just talk and say. According to the language data, non-adult activities include *dirty talking*, *bragging* and *scrounging*. The latter of the three is perhaps the clearest indication of the powerless position of the under-aged. Another characteristic feature seen in boys but rarely mentioned in men is emotionality. Thus, the boy-specific verbs include, for example, *kilkama* 'scream', *kisama* 'shout', *itsitama* 'giggle' *kallistama* 'hug', *pabistama* 'jitter' and *võpatama* 'wince'.

Thus, according to the analysis of interaction verbs, adult men are less frequently depicted as subjects of emotional activities than are women or boys while men are mentioned as subjects of verbal activities as frequently as women and several times more often than boys.

MAN IN PUBLIC AND PRIVATE SPHERES

Men's lower association with the private sphere follows from less importance being attached to their marital status or female connections. Woman-specific adjectival modifiers include *vallaline* 'single', *endine* 'ex-', *praegune* 'current' and üksik 'solitary', but a man's marital status is hardly ever mentioned, i.e. the collocates of *mees* 'man' do not include any man-specific adjectival modifiers referring to marital status. An additional word mostly referring to a woman's family relationship is *uus* 'new', but, as the phrase *uus mees* 'new man' is also frequent, the adjective was not found to be woman-specific. However, an analysis of concordance lines revealed that, with women, *new* mainly refers to her relationship with a man, whereas, with men, the same word is rather associated with professional challenges. In nine cases out of ten, the phrase *uus naine* 'new wife' stands for a new mate (see Example 3), whereas *uus mees* 'new man' is used exactly as frequently in the sense of filling a vacancy (see Example 4).

3. '*Siis sa pead* ***uue naise*** *võtma,' naljatas Tiina toona.* (***Eesti Päevaleht*** 24.10.2000)
'If so, you'll have to get yourself a **new wife**,' Tiina had said jokingly.

4. *Ta värbas sel aastal viis* ***uut meest****.* (***SL Õhtuleht*** 22.09.2007)
He recruited five **new men** this year.

Men's closer association with the public sphere is revealed by the words referring to social status and affluence, which make up a large share of man-specific adjectives. Adjectives such as *tähtis* 'important', *rikas* 'rich', *jõukas* 'prosperous', *rahvusvaheline* 'international' and *vaene* 'poor' are specific to men from the aspects of gender as well as of age. In addition, we find adult men being described as *kuulus* 'famous', *edukas* 'successful', *kodutu* 'homeless', *töötu* 'unemployed' and *töötav* 'working' more often than boys. It is worth noting that the latter word also occurs among the collocates specific to *naine* 'woman; wife', which means that *working* is significantly more frequent among the collocates of *man* than among those of *boy*, but, in the case of the gender-opposed pair *mees-naine* 'man/woman', *töötav* 'working' occurs several times more often with *naine* 'woman; wife'. This is an interesting finding in view of the assumption that, in Estonian culture, where working is an important part of a man's identity, the phrase *töötav mees* 'working man' can be expected to occur as often as or even more frequently than *töötav naine* 'working woman'. A closer analysis of the concordance lines provides an answer to this. With *mees*, in nine cases out of ten, the collocate *töötav* belongs to a three-member phrase where *töötav mees* is modified by a job reference (see Example 5). With *naine*, the collocate *töötav* occurs over three times more frequently, but, in half of the cases, there is no job reference, which means that the phrase *töötav naine* 'working woman' means just an employed woman, i.e. not a housewife (see Example 6). Consequently, women are considered more likely to stay at home, whereas, for men, the employed/stay-at-home opposition is practically inconceivable, which calls for a specification of the man's job or office.

5. *Klienditeenindajana* ***töötav mees*** *ei näe asja tumedates toonides ...* (***Eesti Päevaleht*** 10.11.2004)
The man working in customer service does not look particularly worried ...

6. *Uurimus näitas, et* ***töötavatel naistel*** *oli koduprouadest madalam kolesteroolitase, vererõhk ja kehakaal.* (***Eesti Päevaleht*** 25.01.1999)
The research showed that working women had a lower cholesterol level, blood pressure and body weight than housewives.

No boy-specific adjectives marking social or marital status were revealed. Similarly, according to Baker's findings for English data, men have, through the ages, been described by many words of power and success, such as *rich*, *wealthy*, *grand*, *famous* and *distinguished*, which hardly ever appear in female contexts (Baker 2010: 138). Obviously, the gender-specific adjectives reflect what is considered important for either gender. For a woman, it is her relations with a man – ex-wife, new wife or a single woman. In a man's case, his social status comes to the fore – rich or poor, his role in society. Neither specification works for non-adults.

Although modern women's chances to find employment outside the home are almost as good as men's, the post-industrial stereotypes of man as the breadwinner and of woman as the one responsible for housework are slow to change in Europe (Huntoon 2009: 379). Actually, women's entrance to the labour market does not seem to have brought much change to the division of housework. In Estonia, for example, women do twice as much housework as men (Talves 2011: 103). According to gender monitoring conducted in 2009, Estonian women are mainly responsible for laundry and clothing maintenance, cooking, washing up, tidying up etc., whereas men take care of emergencies, such as repair work and car maintenance (Vainu et al. 2010: 123). This division of labour is well reflected in specific verbs. The only man-specific verb is *ehitama* 'build' (which may mean a home emergency as well as paid employment) while women-specific verbs include *pesema* 'wash; launder', *kuduma* 'knit' and õmblema 'sew'.

In addition to general housework, women with kids usually communicate with kindergarten and school, play with children, help them and provide transport (Vainu et al. 2010: 123). This is reflected in the woman-specific verbs *kasvatama* 'rear' and *toitma* 'feed', as well as in the adjectival modifiers *kasvatav* '(child)rearing' and *kodune* 'stay-at-home'. In addition, woman-specific collocates include the verbs *rasestuma* 'to become pregnant' and *sünnitama* 'give birth', which signify reproductive functions, and the adjectives *rase* 'pregnant', *sünnitav* 'delivering', *(last) ootav* 'expecting (a child)', *viljatu* 'sterile' and *imetav* 'breastfeeding', which are associated with fertility. The analysis shows that, first and foremost, these collocates all have to do with biological sex. Men's most frequent collocates did not include verbs or adjectives associated with biological sex or fertility.

MAN'S APPEARANCE AND PERSONALITY

Descriptions of men's appearance stand out for the use of the adjectives *must* 'black', *turske* 'sturdy' and *verine* 'bloody', which are not found among the collocates of women or boys. Of these three, *verine* 'bloody' belongs to crime statistics, but the former two do not, in most cases at least. The adjective *must* 'black', which mainly signifies racial affiliation, usually occurs in descriptions of musicians, sportsmen, etc. As can be seen from the examples below, the word need not mean the darkest possible skin colour.

> 7. *See on* ***must mees****, pakistanlane, ta ei tohi minna, ära teda sisse lase!* (*Eesti Päevaleht* 03.12.2005)
> This is a black man, a Pakistani, he can't go through, don't let him in!

> 8. *Belgradi jaama ümbruses tiirleb* ***musti mehi*** *nagu kärbseid.* (*Areen* 1998)
> Black men are crowding around the Belgrade railway station like flies.

Although the Estonian word *must* may mean not only 'black' but also 'dirty, unclean' (see EKSS 2009), in the corpus analysed, the collocation *must mees* 'black man' did not reveal any uses of the latter sense. In addition, men and not boys are described in this data as *mustanahaline* 'black-skinned' or *valge* 'white'. Moreover, adult men are sometimes described as *hallipäine* 'grey-haired'. In comparison with women, men's bulk is emphasized by using adjectives such as *turske* 'sturdy', *suur* 'big' and (...kg) *kaaluv* 'weighing (... kg)' while the collocates of *naine* 'woman; wife' include, besides ülekaaluline 'overweight', several words underlying their fragility and vulnerability, such as *habras* 'delicate', *sale* 'slim' and *blond* 'blonde'.

The analysis of gender-specific adjectives showed that attractiveness or its absence is seldom emphasized in men, whereas women are described as being *kaunis* 'beautiful', *ilus* 'pretty', *naiselik* 'womanly' or *inetu* 'plain, ugly' several times more often. The frequent adjectives of attractiveness, such as *ilus* 'pretty', *kaunis* 'beautiful' and *kena* 'nice', constitute less than 1% of all adjectives modifying the word *mees* 'man', in comparison to the over 5% reading for *naine* 'woman' (Piits 2015: 129). Hence, the conclusion that, Estonian descriptions of men's external beauty is an insignificant issue mentioned several times more seldom than in the case of women. However, the word *ilus* 'pretty; handsome' does belong to the 30 most frequent adjectival collocates of both *poiss* 'boy' and *mees* 'man' (ibid.: 124). Paul Baker, in his diachronic study of the collocations of *man*, has found that, compared to English in the previous century, there is

a growing tendency of describing men as *pretty*, *dapper* and *handsome* (Baker 2010: 138), words which directly refer to attractiveness. There are no comparable diachronic studies as yet for Estonian. Nudity is also mentioned rarely with men; the word *paljas* 'naked', for example, occurs four times more often with *naine* 'woman' than with *mees* 'man'. A man is never described as a sexual object; rather, his racial affiliation or bulk is pointed out.

A man's personality, psyche and behaviour are described in much more detail than his physique. The negative traits mentioned include *vihane* 'angry', *karm* 'rough', *kahtlane* 'suspicious' and *vägivaldne* 'violent'. General negative evaluation is expressed by the word *vale* 'wrong'. Positive man-specific personality traits and general evaluatives from both gender and age aspects are *aus* 'honest', õige 'right', *vajalik* 'necessary', *mõistlik* 'reasonable' and *mõnus* 'cool'. The meaning of the adjective *aus* 'honest' differs depending on whether it refers to a man or woman. *Aus mees* signifies a morally virtuous person (see example 9) while the few cases of *aus naine* refer rather to a sexually virtuous woman with strict moral standards (see Example 10):

> 9. *Ta on väga **aus mees**. Ta ütles, et ei taha valetada.* (*Eesti Päevaleht* 18.03.2006)
> He is a most honest man. He says he doesn't want to lie.

> 10. *Ma ei kujuta üldse ette, kuidas üks **aus naine** suudab rõõmu tunda selliste võltside vahendite abil saavutatud seksuaalvahekorrast!* (*Eesti Päevaleht* 10.03.2001)
> I can't imagine how an honest woman could enjoy sexual intercourse achieved by such fake means!

Age-specific adjectives describing *man* also include *tark* 'clever; wise', *tõeline* 'real', *kange* 'efficient' and õnnelik 'happy'. Thus, *wisdom*, *efficiency* and *happiness* are associated with adulthood while persons under age evoke such attributes as *talented*, *naughty* and *cute*. The words *kõva* 'tough', *tore* 'nice' and *asjalik* 'no-nonsense', however, are used independently of the age of the male person.

MAN AS AN ADULT

The biggest group of age-specific adjectives for men defines their age. Most of the adjectives end in *aastane* '... years old' while *keskealine* 'middle-aged', *eakas* 'elderly' and *vana* 'old' specify age less definitely and more subjectively. Although *mees* is applicable to all age groups, exact specifications invariably refer to persons aged 18 or older, whereas the exact specifications of *poiss* refer to boys under 18. '18-year-old' is the only exact age attribute used with *mees* as well as with *poiss*; in all other cases where *aastane* is used, a clear difference is made.

Power in the family largely depends on who decides how family money should be used. According to resource theory, the power balance tilts in favour of the party who contributes more to the family's income (Talves 2011: 111). The adjectives referring to man's affluence show that the husband is mostly assumed to be the main contributor while power lines need not run between different sexes at all. Man occurs as the subject of the verb *maksma* 'pay' several times more often than is the case for boy. Lack of power in the case of non-adults is reflected in the verbs *kuuletuma* 'obey', *imetlema* 'admire' and *järgima* 'follow' collocating with *poiss* 'boy'. Violence and criminal behaviour have also been interpreted as demonstrations of power. However, here, my method fails to reveal major differences because various aggressive activities are mentioned in connection with men as well as boys, although the activities differ.

A comparison with boy-specific verbs enables the conclusion that *learning* and *acquisition*, as well as certain creative activities like *handicrafts*, *drawing*, *singing* and *dancing*, do not belong to adult activities.

CONCLUSION

This article is focused on differences in Estonian media descriptions of men, women and boys as reflected in specific adjectives and verbs. It is true that there are also great similarities in the collocational behaviour of the words in question. For example, comparing the lists of the 30 most frequent verbs belonging to the words *mees* 'man', *naine* 'woman' and *poiss* 'boy', we find that two thirds of them coincide (Piits 2015: 106–107). And yet, the differences are eloquent enough to reveal how, in comparison with boys and women, the Estonian-language media describes men's appearance and personality, their peculiarities of verbal expression and their activities in the private and public spheres.

According to my analysis of interaction verbs, adult men are less frequently presented as subjects of various emotional activities than are women or boys

while man functions as the subject of various verbal activities as often as woman and several times more often than boy. Man-specific interaction implies a power position; man *orders* and *commands*, *contends* and *blames*.

Violence and criminal behaviour have also been interpreted as a demonstration of power. Descriptions of men are conspicuous for adjectives associated with criminal or deviant behaviour, especially in comparison with women, while boys also engage in aggressive and deviant activities.

Unlike woman, man is not described as a sex object; instead, his physical aspect is marked by mentioning his race and bulk. Man's physique is described in far less detail than his personality, psyche and behaviour. Man is usually seen as active in public life, and his private sphere is less in focus. His social status is important, his being rich or poor, as is his role in society.

ACKNOWLEDGEMENTS

This work was supported by institutional research funding IUT 35-1 of the Estonian Ministry of Education and Research, TK145 "Centre of Excellence in Estonian Studies – CEES", and the governmental basic financing of the Institute of the Estonian Language from the Estonian Ministry of Education and Research.

REFERENCES

Ahven, Andri & Kruusmaa, Kätlin-Chris 2013. *Vangistus ja kriminaalhooldus: Kuritegevus Eestis 2012*. [Imprisonment and Probation: Criminal Activity in Estonia 2012.] Tallinn: Justiitsministeerium.

Baker, Paul 2010. Will Ms Ever Be as Frequent as Mr? A Corpus-Based Comparison of Gendered Terms across Four Diachronic Corpora of British English. *Gender and Language,* Vol. 4, No. 1, pp. 125–149. DOI: 10.1558/genl.v4i1.125.

Biber, Douglas & Conrad, Susan & Reppen, Randi 1998. *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge: Cambridge University Press.

Butler, Judith 1990. *Gender Trouble: Feminism and the Subversion of Identity.* New York & London: Routledge.

Caldas-Coulthard, Carmen Rosa & Moon, Rosamund 2010. 'Curvy, Hunky, Kinky': Using Corpora as Tools for Critical Analysis. *Discourse & Society*, Vol. 21, No. 2, pp. 99–133. DOI:10.1177/0957926509353843.

EKSS 2009 = *Eesti keele seletav sõnaraamat I–VI.* [The Explanatory Dictionary of Estonian I–VI.] Tallinn: Eesti Keele Sihtasutus.

Huntoon, Alishia 2009. Gender Stereotypes. In: Jodi O'Brien (ed.) *Encyclopedia of Gender and Society* 2. California: SAGE Publications. http://dx.doi.org/10.4135/9781412964517.n189.

Hyde, Janet Shibley 2005. The Gender Similarities Hypothesis. *American Psychologist*, Vol. 60, No. 6, pp. 581–592. http://dx.doi.org/10.1037/0003-066X.60.6.581.

Kallas, Jelena 2013. *Eesti keele sisusõnade süntagmaatilised suhted korpus- ja* õppeleksikograafias. [Syntagmatic Relationships of Estonian Content Words in Corpus and Pedagogical Lexicography.] Tallinn: Tallinna Ülikool.

Kilgarriff, Adam & Rychly, Pavel & Smrz, Pavel & Tugwell, David 2004. The Sketch Engine. In: G. Williams & S. Vessier (eds.) *Proceedings of the 11th EURALEX International Congress*. Lorient: Université de Bretagne Sud, pp. 105–115. Available at http://www.euralex.org/proceedings-toc/euralex_2004/, last accessed on May 23, 2106.

Kilgarriff, Adam & Kovar, Vojtech & Krek, Simon & Srdanovic, Irena & Tiberius, Carole 2010. A quantitative evaluation of word sketches. *Proceedings Euralex*. Leeuwarden, The Netherlands: Euralex, pp. 372–379. Available at http://www.euralex.org/proceedings-toc/euralex_2010/, last accessed on May 23, 2016.

Liljeström, Marianne 2003. Soosüsteem. [Gender System.] In: A. Koivunen & M. Liljeström (eds.) *Võtmesõnad*. [Keywords.] Tallinn: Eesti Keele Sihtasutus.

Lindsalu, Pilleriin 2012. Kuritegevus Eestis 2011. aastal: kokkuvõte. [Crime in Estonia 2011: Summary.] *Kuritegevus Eestis 2011.* [Crime in Estonia 2011.] Tallinn: Justiitsministeerium.

Nuolijärvi, Pirkko & Tiittula, Liisa 2000. *Televisiokeskustelun näyttämöllä: televisio-institutionaalisuus suomalaisessa ja saksalaisessa keskustelukulttuurissa.* [Television Discussion Stage: TV-institution in Finnish and German Conversation Culture.] Helsinki: Suomalaisen Kirjallisuuden Seura.

Panther, Klaus-Uwe & Thornburg, Linda L. 2012. Antonymy in Language Structure and Use. In: Mario Brdar & Milena Žic Fuchs & Ida Raffaelli (eds.) *Cognitive Linguistics between Universality and Variation*. Newcastle, UK: Cambridge Scholars, pp. 161–188.

Pearce, Michael 2008. Investigating the Collocational Behaviour of MAN and WOMAN in the BNC Using Sketch Engine. *Corpora*, Vol. 3, No. 1, pp. 1–29. DOI: 10.3366/E174950320800004X.

Piits; Liisi 2015. *Sagedamate inimest tähistavate sõnade kollokatsioonid eesti keeles.* [Collocations of the Most Frequent Estonian Words for 'Human Being'.] Tartu: Tartu Ülikooli Kirjastus.

Pilvre, Barbi 2011a. Meedia ja sugu. [Media and Gender.] In: Raili Marling (ed.) *Sissejuhatus soouuringutesse.* [Introduction to Gender Studies.] Tartu: Tartu Ülikooli Kirjastus.

Pilvre, Barbi 2011b. *Naiste meediarepresentatsioon Eesti ajakirjanduskultuuri ja* ühiskonna *kontekstis.* [Media Representation of Women in the Context of Estonian Journalist Culture and Society.] Tartu: Tartu Ülikooli Kirjastus.

Talves, Kairi 2011. Soosotsioloogia. [Sociology of Gender.] In: Raili Marling (ed.) *Sissejuhatus soouuringutesse*. [Introduction to Gender Studies.] Tartu: Tartu Ülikooli Kirjastus.

Vainu, Vaike & Järviste, Liina & Biin, Helen 2010. *Soolise võrdõiguslikkuse monitooring 2009*. [Studies of Gender Equalities 2009.] Sotsiaalministeeriumi toimetised 1. Tallinn: AS Ecoprint.

http://dx.doi.org/10.7592/FEJF2016.64.voolaid

CHILDREN'S FUNNY REMARKS IN THE FIELD OF LINGUISTIC HUMOUR THEORY

Piret Voolaid

Abstract: The article analyses, from folkloristic and humour theoretical aspects, humorous material of children's remarks collected during the all-Estonian kindergarten folklore collection campaign held from October 2010 to January 2011.

The main focus is on this subtype of jokes as they appear in kindergarten environment and from the point of view of kindergarten teachers. The material is divided into two groups: 1) spontaneous sayings, recorded during daily activities and interaction; 2) answers to the teacher's questions, guided by her interest (the teacher may have recorded discussions on a given topic). The article aims to investigate the utterances that teachers have perceived as funny or worth recording and to analyse the theoretical mechanisms of humour they are based on.

Keywords: all-Estonian kindergarten folklore collection campaign, child language, child lore, children's funny remarks, linguistic theory of humour

We have all heard, some of us more than others, the candid and direct remarks children make that put a humorous twist on reality. Parents who witness their children growing up probably hear these remarks more often, but pre-school teachers who spend even more time with children during their waking hours than their parents also hear them a great deal. The widespread custom of writing down what children say can be regarded as part of family and pre-school lore. Children's remarks are often circulated in video, audio, and social media as a separate form of humour. Discussions of adult topics from a child's point of view as well as the imaginative linguistic creativity of children come across as sincere, genuine, and often funny. This sincere and genuine way of speaking has also brought the word *lapsesuu* ('child's mouth') into the Estonian language, which is used figuratively to characterise someone who speaks as frankly as a child (EKSS 2009: 55). Sounding childish can sometimes be an intentional rhetorical method, similar to the way politicians sometimes speak.

Writing down the funny things children say is a popular tradition in modern written culture, but Estonian folklorists have not carried out any detailed research on the subject to date. The main reason is that such funny remarks are associated with a specific child at the moment they are uttered, so they

are the creations of an author, which is something that only began to receive attention in folklore a couple of decades ago. The systematic collection of child-lore in Estonia began in the 1920s,[1] but a catalyst in research occurred after a nationwide competition in collecting school lore that took place in 1992 (see, e.g., Kõiva 1995)[2]. However, the rich world of children's lore has many facets that are still waiting to be analysed.

Since the autumn of 2010, I have had the opportunity to participate in the organisation of a nationwide series of Estonian language training seminars initiated by the Ministry of Education and Research as a coordinator from the Estonian Literary Museum. Writers, researchers, museum workers, and pre-school educators traditionally speak at the two-day seminars now held in Tartu every year. The topics discussed include the meaning and values of language use in society and the options for using cultural heritage and folklore (language also represents the building blocks used to create folklore) in language training. At the first training seminar, I led a workshop about children's lore and the work we did prior to the seminar gave us the idea to announce a nationwide competition in collecting pre-school lore, which would be organised jointly by the Department of Folkloristics of the Estonian Literary Museum and the Estonian Folklore Archives.[3] One of the responses we received to the in-depth questionnaire we sent out was over 100 pages of the cool expressions and funny remarks of children, which are the source material for this research.

The purpose of the article is to take a look at the material we received, the main research questions being: 1) how can children's jokes be defined; what is the bigger picture revealed by the phenomenon; 2) what has been perceived funny by teachers and what have they decided to write down; and 3) on which mechanisms of the theory of humour is children's humour based and which principles can be used to systematise the children's jokes that have been written down.

INTERDISCIPLINARY RESEARCH OF MATERIAL

Researchers trying to define and analyse children's jokes immediately find themselves in an interdisciplinary field. There is no doubt that the remarks that are on the borderline of language and folklore are part of a broader children's language, which is why the aspects related to the development of children's language skills and speech must be considered when their jokes are studied.

In terms of folklore, children's jokes fall into the category of funny things that have happened in real life and (language specific) stories based on humorous life events (see Hiiemäe 2014: 845). Mall Hiiemäe (ibid.) has written that

collecting (humorous) folk tales based on individuals and influenced by real life started in the 1930s, but an earlier scientific approach and the more targeted definition of such material emerged in Estonian folklore in the 1960s. The term *"lapsesuu"* ('child's mouth'), which is the precursor of the term *"lapsesuunaljad"* (literally 'jokes from a child's mouth'; 'children's jokes'), was already used in the introduction to the first volume titled *Periods of Human Life* of the publication *Funny Estonian Folk Tales*, written by Rudolf Põldmäe in 1941:

> *A person becomes the butt of jokes as soon as he or she is born. Experienced observers laugh at the ignorance of those who know nothing about the emergence and development of life. There is also no way of avoiding laughter during the trials and tribulations of a child's spiritual development. This creates the so-called* ***"lapsesuu" tales*** *('funny folk tales from children'), which have been actively published by our former newspapers and family magazines, and which have also made their mark on funny folk tales. Types that are too literary have been excluded from this, although the issue of sources is far from resolved. On the other hand, the logic and tales of children are based on true stories that have not yet acquired the traditionalism of folklore. However, there are still many ancient motifs in this group, which have lost their sharp corners as they have become widely spread and have acquired a certain artistic form.* (Põldmäe 1941: 8)

Therefore, Põldmäe documented the phenomenon and also raised the issue of authenticity and secondary literature, which was important from the viewpoint of folkloristics of the time. Also, he already pointed out the influence of media on the distribution of children's jokes. The jokes themselves (no. 15–96) have been divided into sections in the book: *Children's Stupidity and Mishaps*; *Children's Candour*; *Children's Laziness, Stubbornness and Malice*; *Children Acting as Adults, Children's Wisdom and Shrewdness*. Sigrid Schmidt (2005: 257) highlighted three types of children's jokes: 1) jokes whose heroes are usually stereotyped figures, which is the category in which she places the majority of folklore jokes; 2) jokes told by children to children and which also confront child heroes and adults; 3) actual remarks by children usually addressed to adults that adults regard as amusing though the children are serious. Whilst the children's jokes given by Põldmäe in his *Funny Estonian Folk Tales* mostly represent the first sub-type, the jokes written down by pre-school teachers mainly belong to the latter group. The difference from anecdotes lies in the fact that the remarks are usually not funny for children themselves, but adults see the jokes that have emerged in everyday situations as unintentional humour (see Martin 2007: 1415) or accidental humour (Nilsen & Nilsen 2000: 6–9).

Similar to the collection and research of childlore, the collecting and researching of children's language in Estonia also started in the 1920s. The first invitation to start writing down remarks made by children was published by Julius Mägiste in his article *A Few Words about Child Language*, which was published in the first issue of the *Eesti Keel* (Estonian Language) magazine in 1924 (Mägiste 1924). Andrus Saareste and Paul Ariste also studied child language in the 1930s (Argus 2003). The level of collecting and researching child language in present-day Estonia is professionally high and the representative corpus created on the basis of recordings of everyday speech (the first data were added in 1998 (Argus 2007)) has been logged in the international CHILDES system (Child Language Data Exchange System), which was developed in order to provide a common basis for transcribing, processing, comparing and sharing the language material collected by different researchers. The linguistic goal of such corpora is to study the problems associated with the acquisition of a first and second language, bilingualism, and various clinical problems. Several studies of how a small child learns his or her native language have been published on the basis of the corpus. In purely linguistic terms, the sub-corpus of the article also reflects the essence of the language acquisition process, the logic behind child language (repeated slips of the tongue) and the way they understand language. Against the background of said material, it also became evident that speech therapists in pre-schools write down the witty remarks made by children out of professional interest in order to analyse technical errors in speech.

Several databases of child language outside Estonia can also be used as examples. The corpus of remarks characteristic of Russian child language prepared by Russian linguist Vera Kharchenko (2012) was created as a result of a project that lasted from 2005 to 2011. The material was collected during the extensive observation of the everyday language used by the researcher's two grandsons (born in 2003 and 2006). Kharchenko calls the corpus she created an alternative dictionary genre, which is another arsenal of the specific features of child language, which allows researchers to identify and study child language from broad interdisciplinary aspects. Kharchenko's longer-term goal is to combine her work with previously collected material to prepare a multivolume dictionary of Russian child language. The author herself proceeded from the following technical aspects in her work: the fact of word and form creation, congruency errors, wrong word stress, child's questions, child's discussions (with the emphasis on the child's opinions and beliefs), copying adults (Kharchenko 2012: 15).

The entire discourse of children's jokes must also be approached from the pedagogical angle. Once again, we are facing the fact that the humour arises from the position of adults, who keep passing on the joke and may even start using the funny linguistic form themselves. However, educationists have em-

phasised that it is pedagogically wrong to repeat a child's speech using the words they have said incorrectly, even if they may seem funny and interesting to adults, because there is nothing interesting in them for children – they do not realise that they say them differently from adults; laughing and making fun of them will hurt and confuse them (Kraav 2007: 75). This shows that different disciplines dealing with the same research material emphasise different things.

ASPECTS OF HUMOUR THEORY

The way children experience and express things is usually turned into a story by someone else, usually a family member – parent or grandparent, older brother or sister, child-minder, pre-school teacher, etc. – who happened to witness the event. What makes these experiences worth recording is their different logic, mistakes, developments that seem interesting and emotionally valuable. Children's jokes are a part of everyday speech. They usually come across as entertaining, spontaneous, and mundane situational humour that children themselves ordinarily do not see as funny. Inside lore groups, the tales usually relate to specific children and events, but good story lines may spread outside the specific community (e.g. via collections of jokes or by being told to others as anecdotes) as anonymous jokes.

Those who collect children's remarks have called them expressions of the genius and creativity of children's way of thinking (Chukovsky 2001). The cultural and historical joke theories aimed at social aspects describe humour as a phenomenon that reflects the society and can be used to discover concealed views of what is going on in society, the contradictions or problems (Laineste 2003: 798). This means that a question we could ask in this study is to what extent is the surrounding reality reflected through the eyes of children and what can be said about the lives of children in a broader sense based on their remarks.

In a larger social and cultural context, the research of jokes based on the remarks of children is compatible with the changes in the approaches to childhood that have occurred in recent decades, in moving the emphasis from the traditional towards a more sociological approach. Due to their immaturity and incompetence, children have traditionally been regarded as beings that depend on adults, but the viewpoint of childhood sociology is child-centred – a child is an active member of society who is given a greater say and whose opinion is taken into account (Prout & James 2005: 59). Giving attention to what children say attributes importance to the child as a personality and their opinions; it makes them heard. As the material becomes more folkloric, funny remarks have the potential to become group or family lore, which may remain part of

the community's language for years. The gradual disappearance of the original context and personality strengthens the importance of the funny saying, which may become rooted in community slang as an especially short form of a word, or stay in circulation as a longer narrative. Funny stories may become more topical as the child grows and his or her reaction to it may be ambivalent. Jokes may become tools that support the child's identity, helping them develop their self-image and describing the child in a period of time they do not remember ("I said it like this."). These recorded remarks therefore enrich the child's knowledge of self in social situations and relationships, supporting memories of their personal past. However, it is possible that telling jokes about the child may cause embarrassment to him or her once they reach a certain age.

So-called cognitive humour theories gained considerable ground in humour studies in the 1960s and 1970s. They regard humour as a cognitive experience and find that there is some kind of objective 'real' contradiction or incongruence in a funny (verbal or non-verbal) object, which creates a moment of surprise and comes across as funny (Krikmann 2004: 6). Based on the material in hand, it is the humour in the remark that makes an event worthy of recording and sharing with others; the moment of surprise that makes a remark funny lies in the child's unexpected logic, the way in which they understand or misunderstand things, slips of the tongue that are funny in certain situations or *per se*, which are emotionally worthy of being written down and passed on to others. Contradictions, exaggerations, subconscious mistakes and children's genuine and sincere candour often differ from the way adults speak and surprise the listener.

Alleen Pace Nilsen and Don L. F. Nilsen (2000: 6) have divided accidental humour into linguistic (verbal) and physical. Child humour as verbal source material falls into the category of linguistic humour, which is why it has a secure place in linguistic theories of humour. In addition to children's remarks, such accidental humour also covers the funny remarks made by reporters (sports commentators), funny quotes from the works of students (e.g. essays, research papers) and many others that all have relatively universal bases. Child lore can also be adequately researched in light of the General Theory of Verbal Humour (GTVH) of Salvatore Attardo and Victor Raskin (1991), which generally lies in the hierarchic representation model of six knowledge resources. In brief, the levels of knowledge resources are: 1) script opposition (SO), which was taken over from Raskin's (1985) earlier semantic theory of humour and which states that a joke must be compatible with two different scripts and these in turn must be in opposition to each other to a certain extent; 2) logical mechanism (LM), which when breached causes discrepancies and false analogies that evoke humour; 3) situation (SI), which forms the contextual foundation of the joke and includes activities, participants, objects, etc.; 4) target (TA) or the butt

of the joke; 5) narrative strategy (NS) or the genre in which the joke works (e.g. anecdote, proverb, riddle), in this case longer or shorter texts that briefly describe activities and are presented as monologues or dialogues where the main emphasis is on the child's remark; 6) language (LA) or the actual lexical, syntactic, phonological, and other choices necessary for the emergence of humour (Attardo 2010 [2008]: 108; Krikmann 2004: 54–63). The theory can be applied separately to all humorous texts, but due to the source material of this article I will focus my analysis on the linguistic level relevant to the material, which has been considered the base level influenced by the others. However, I will still point out the connections between the material and all other levels.

MATERIAL, METHOD OF COLLECTION, AND ANALYSIS

Children's jokes in the context of pre-schools and the viewpoint of pre-school teachers are in the foreground in the article based on the material obtained from the competition in collecting child lore. The survey plan used for the child lore collection competition of 2011, where participants were asked to give their answers in free format, covered three topics: 1) Festive occasions and parties; 2) Games; 3) Tales and remarks. There was a separate point C under the third topic: Please observe children for some time and write down any funny remarks they make. This was answered by 45 pre-school employees (teachers, speech therapists) from 34 Estonian-language pre-schools nationwide. The material received consisted of 100 pages of material and 880 children's jokes written down by adults. Shorter responses contained single remarks written down as a result of brief observation, while longer responses contained over 120 texts (e.g. the 129 written recordings from 2001–2010 sent by the teachers of Tartu Kannikese pre-school).[4]

The collection competition confirmed the fact that teachers in many pre-school groups have initiated the tradition of writing down the cool and witty remarks of children. Some people have found ways of preparing and publishing collections of these remarks, creating group lore that will always remind the people concerned of the time spent in pre-school. Officially, such material belongs in the child's learning or development portfolio (development folder), which is kept for each child during the time they attend pre-school and which plays an important role in the process of assessing the child's development. Teachers of the Helika pre-school in Tartu mention the emotional value of the collected material in their cover letter, saying that it makes a great gift. The teachers and last-year pre-school children prepared illustrated *Lapsesuu* (Child's Mouth)

books, which were given as gifts on the nursery school's birthday or to children leaving pre-school for primary school (ERA, DK 40, 202 (3.C)).

There have also been other child lore collection initiatives and teachers gladly take part in them. The aforementioned cover letter says that two teachers of Helika pre-school in Tartu won the joke competition organised by children's magazine *Täheke* (Little Star) in 2009 with children's jokes, after which the jokes were published in both *Täheke* and the compilation *Kükitav mannatera* (Squatting Manna Grain) (Martson 2010).

A pre-school teacher from Hargla said that the remarks made by children are recorded in a separate notebook to make sure "they are not forgotten" and read out to the children every April Fool's Day, April 1 (ERA, DK 40, 180 (3.C)).

Pre-school teachers generally added no information about naming the phenomenon or terminology. As respondents were asked to write down funny remarks made by children, there was no need for them to use terminology in respect of the phenomenon.

Written recordings are generally divided into two: 1) remarks made during incidental everyday activities and communication; and 2) remarks received in response to targeted questions asked by teachers and their expectations (the teacher has recorded a discussion of a given subject, e.g. the meaning of happiness, what children want to become when they grow up, etc.). In terms of form (which constitutes the narrative strategy in the aforementioned GTVH), written recordings may only be short monologues of anonymous children (e.g. a noteworthy word used by a child with the meaning presented by the teacher) without any description of the external context or situation:

1. A lexical unit to which the meaning of a word or verbal combination has been added.

   Müri oli – äike müristas. (There was *müri* – a noun made up from the verb *müristama*, to thunder) (ERA, DK 40, 234 (3.C))

2. The remark is presented as direct speech of the child, preceded by the child's first name or first name and age:

   Tormi: "Minul on need täitsa uued peksapüksid (teksapüksid)" ("I have these new beat pants (jeans)") (peksapüksid (peksma – beat, püksid – pants, trousers) instead of *teksapüksid* (jeans)) (Sept 2010) (ERA, DK 40, 235 (3.C))

   Raio (3): "If I had one leg, I'd know exactly where to put the boot!" (ERA, DK 40, 258 (3.C))

3. Often, teachers have written down dialogues between themselves and a child, or between two children, where the joke is evoked by the child's response to the teacher's question or a reaction to some activity:

   The teacher asks Markus Sander: "May I help you to peel this mandarin?" Markus Sander asks back: "Do you want to eat the peels?" (ERA, DK 40, 126 (3.C))

4. Less often, teachers present narratives with longer contextual descriptions, where the child's remark provides the comical punch line.

   A 4-year-old boy was watching a burning candle and, with my permission, started to put it out with a snuffer. He pressed it down too hard on the candle, making hot candle wax drip on the floor and congeal. I said: "Now that's trouble." The boy responded: "That's not trouble; that's an occupational accident". (ERA, DK 40, 118 (3.C))

This is a case of short folklore, where the narrative description of activity is brief or missing altogether and the child's remarks provide the punch line that creates the comical effect.

In some cases, teachers have added cover letters. A teacher from Priisle preschool in Tallinn has mentioned the importance of the situational context and emphasised that "often, these jokes are only funny in the given situation. The child's facial expressions, body language, etc., also add to this" (ERA, DK 40, 222 (3.C)).

It must be said that the context, the situation (level 3 of the GVTH), the syncretic and multimodal facets of the material are briefly described on the basis of the recorded remarks. However, these details really are important for the emergence of the comical effect, e.g. in the next example, a child sitting with his leg crossed over his knee comes across as comical to the teachers and is an important detail in the emergence of humour:

> *Raul (4) sits with the teachers in the recreation room, with his leg crossed over his knee, and declares: "Imavere pre-school is a really good pre-school!"* (ERA, DK 40, 358)

There are many texts whereby the context must be explained by the teacher before its final meaning can be understood:

> *Agnes (4) calls out to the teacher: "Tule ruttu, Henri hakkab siin majandust tegema!" ("Come quickly, Henri is starting to make economy here!")* (*majandus* – economy, *pahandus* – mess). (ERA, DK 40, 329)

This survey is a primary attempt to categorise material. The material can be divided into a system based on content and subject matter, which I have used in the gift book *Meie armastuse emmed* (Our Beloved Mums) commissioned by Ajakirjade Kirjastus (Voolaid 2012): e.g. family (mums and dads, grandmas and grandpas), nature (plants, animals, birds), people's appearance (beauty), festive occasions, everyday life (eating, getting dressed, playing), abstract phenomena (luck, accident, wealth, etc.). In the case of classification by content and subject matter, a text may belong to several categories at the same time. In her study, Sigrid Schmidt (2005) highlighted four main categories of children's remarks: jokes caused by insufficient language skills, limited knowledge of the adult world, questions, witty answers and reactions. Schmidt's categories are presented from different bases or overlapping: the last to seem to be rather formal, e.g. a child's question or witty answer/reaction may be funny because of insufficient language skills or limited life experience, and also arise from limited knowledge of adult life. According to the GTVH, jokes arising from the insufficient language skills of a child work primarily on linguistic resource; a child's limited knowledge of the adult world may also manifest itself at the level of language, but at the same time reflect the child's understanding of their surroundings and reveal references to the surrounding context, the sociocultural background.

Child humour based on linguistic resource or language humour

The linguistic level mentioned in the GTVH seems to be the most important factor in the analysis of children's jokes and is primarily expressed here in pronunciation errors, mix-ups by the speaker, such as Freudian slips, malapropisms (the act of using an incorrect word in place of one that is similar in pronunciation, e.g. using similar-sounding foreign words), spoonerisms (a verbal error in which a speaker accidentally transposes the initial sounds or letters of two or more words), etc. One-third of the remarks written down by pre-school teachers are based on the level of word formation, and they create a source of humour based on wordplay in the classic sense. According to Salvatore Attardo, linguistic jokes occur technically in ordinary speech as wordplay, but they differ in their nature from intentional puns, which are also considered linguistic jokes (similar to meta- and poetic language) (see Attardo 1994). Attardo (ibid.) distinguishes four categories of wordplay with similar words: paronyms, homonyms, homographs, and homophones. Paronyms are words that are similar in their phonemic composition (sound), but that are not identically spelt and have different meanings. Homonyms are similar in their phonemic and graphemic

composition, divided into homographs – words that share the same written form as another word but have a different meaning (*row* : *row*), and homophones – words that are pronounced the same as other words but differ in meaning, and may differ in spelling (*mourning* : *morning*).

One of the main causes of linguistic jokes is the use of a word with similar pronunciation in the wrong meaning, i.e. the lexical intersection of the joke lies mainly in similar words or homonyms and paronyms. Children can mix up words that overlap partly in the morphemic composition. The child uses his or her thoughts, using a word that they have heard and that they associate with the word they are trying to say. Such mistakes sound funny, as they make the listener/reader mentally join two scripts – the one that was said with the one that was meant or intended to be said (see also Nilsen & Nilsen 2000: 8). Of course, one has to keep in mind that we can usually not speak of intentional linguistic 'comedy' in the case of children, which is why the remarks are better described with neutral linguistic terms (metathesis, paronymy, homonymy with its subtypes homography and homophony) instead of intentional methods of humour (wordplay, such as spoonerism, malapropism, Freudian slip, paranomasia/paranomastic image).

At the lexical level, the replacement of letters in a word is enough to lead to a joke:

> *Janek (6): "Marjo, kas sul* ***mokamütsi*** *on?"* (*"Marjo, do you have a* ***lip*** *cap?"*; *mokamüts* – lip cap; *nokamüts* – billed cap) (ERA, DK 40, 358); there is *m* alliteration: *mokamüts* instead of *nokamüts*.

A grammatical slip of the tongue may also make a word humorous, e.g. choosing the wrong thematic vowel or consonant gradation that the child is unable to use normatively:

> *We're driving past the Anne rowing canal and 6-year-old Ander says: "Lähme homme siia* ***kanalasse*** *ujuma"* (*"Let's go swimming in the* ***chicken farm*** *tomorrow"*) (ERA, DK 40, 328); (*kanal* : *kanalisse* – canal; *kanala : kanalasse* – chicken farm).
>
> *5-year-old Sven Erik heard his teacher and mum speaking about blood sausages and found that:* ***"Seasid*** *tapetakse"* (*"Pigs are killed"*, the correct form in Estonian would be *"sigu"*). (ERA, DK 40, 331)

A typical example of a slip of the tongue is the metathesis in the expression *kõrge kontsaga kingad* or *kontsakingad* ('high-heeled shoes' or 'heeled shoes'), whereby teachers have written down mistakes on four occasions (and interestingly, all the mistakes have been made by girls): *kintsakongad* (ERA, DK 40,

260 (3.C)), *konksa kingad* (ERA, DK 40, 343); morphological derivation with the same expression has also been recorded: *ilusad kontsatatud kingad* (ERA, DK 40, 301), *kontsad* (ERA, DK 40, 304). The material contains other metathetic examples at the level of words, which in the theory of humour are treated as spoonerisms:

> *Joosep (4): "Teacher! We went to the hallway!* ***SAJALA****!!!"* (should be *salaja* – secretly). (ERA, DK 40, 257)

The following jokes, for example, are based on paronomastic formation:

> *Gerdo tells the girls: "Ma olen saanud kunagi* ***pasteedi*** *peal hüpata!"* (*"I got to jump on* ***pâté*** *once!"*) (ERA, DK 40, 357); (*pasteet* – pâté; *batuut* – trampoline).

> *Miiu (4): "Minu lemmikloom on* ***amsterdam"*** *("My favourite animal is* ***amsterdam****!"*) (*Amsterdam* instead of *hamster*). (ERA, DK 40, 203 (3.C))

> *Children are eating at the table. Mehis looks out the window at the swirling snow and says: "****Purskab*** *väljas!" ("It's* ***bursting*** *outside!")* (*purskab* – bursting; *tuiskab* – snowing heavily). (ERA, DK 40, 354)

The language skills of a child in pre-school are still developing. They often do not have enough words to express themselves, so they have to invent inflections themselves. These are characterised by playfulness and creativity that have their own logic. Children have a command of various mechanisms. A typical one is blending several words or contamination, e.g. the aforementioned *kontsad*; teachers have twice recorded the use of *pillima* (to cry) to say *pilli mängima* (to play a musical instrument) (ERA, DK 40, 83), and *klaverit mängima* (play the piano) (ERA, DK 40, 180 (3.C)). Morphological derivation, e.g. deriving a verb from a noun, is also characteristic of child language:

> *Õnne Liis goes to the new cleaning corner in the room: "Teen lapi märjaks ja lähen* ***lappima****!"* (*"I'll make the rag wet and go* ***ragging****!"*) (*lapp* – rag, cloth, patch; *lappima* – to mend something by putting a patch on it). (ERA, DK 40, 354)

> *Eliise (2) shows her polished nails: "Näe, vaata – Mann* ***küüsis****!" ("Look – Mann* ***nailed!****")* (*küüs* – nail, *küüsima* does not exist). (ERA, DK 40, 358)

> *Children are sewing. A boy says: "Tüdrukutel tuleb* ***õnelamine*** *paremini välja*. (*"Girls are better at* ***stinging"***) *(õnelama* – to sting; *õnel* – needle). *This is not man's work".* (ERA, DK 40, 170 (3.C))

The mechanism also works in the opposite direction – children derive nouns from verbs:

> *An electrician is working in the room, changing a switch on the wall. Mikael (5): "Kas sa paned meile uue* ***vajutela****?"* (*"Are you installing a new* ***presser****?"*) (*vajutama* – to press; *lüliti* – switch). (ERA, DK 40, 302)

> *Kirke-Liisa: "Gerdo läheb uisutama* ***uisulasse****!"* (*"Gerdo will go skating in the* ***skatery****!"*) (*uisutama* – to skate). (ERA, DK 40, 357)

Children can be very prolific and creative when forming compound words:

> *Children are looking at photos of Estonian celebrities. The teacher shows them a photo of the ballerina Kaie Kõrb and asks: "Do you know who Kaie Kõrb is?" Gerly:* ***"Ilukeerutaja!"*** (***"Figure spinner!"***). (ERA, DK 40, 356)

The material contains original creations arising from the individual logic of children, who have their own etymology:

> *Annabel, two, calls the checkout lady in the shop* ***"maksma tädi"*** (*'pay auntie'*), *because her mum always says in the shop that now we have to pay auntie for the goods.* (ERA, DK 40, 330)

Homonymic creation of entirely new words is also common:

> *"Nüüd ma tean, pull on isa, tal ei ole* ***taguraid****!"* (*"Now I know, a bull is male as he has no udders!"*) (*Tagurad* is used instead of *udarad*). (ERA, DK 40, 315)

> *"Ema, vaata,* ***kapsutaja*** *tuleb!"* (*"Look, mum, the teacher is coming!"*) (*kapsutaja* is used instead of *kasvataja*). (ERA, DK 40, 315)

It is known that the teacher who wrote down the last example uses the word *kapsutaja* as her username in social media and on Skype. In terms of folkloristics and linguistics, it is interesting to note that a child's original (incorrect) linguistic creations can have an emotional effect on an adult's correct language and the wrong form of a word may be used as group slang in a community for years.

The majority of remarks suggest imitation of the correct words heard from adults and repetition of expressions. It feels funny for an adult when children start using the expressions they have heard and place themselves in the world of adults, e.g. use foreign words or opinions that are not age-appropriate. Children's remarks, which are obviously based on a previous description of a situation by an adult, are testament to the connection between the linguistic and contextual levels (manner of speaking and stylistic choices):

> *Melinda (3): "An old hag sold Salme* [grandma] *putrid fish at the market!"* (What vocabulary at the age of three!) (ERA, DK 40, 257)

Unexpected style or the use of a figure of speech may also be funny. For example, the use of slang may be unexpected:

> *Raul is dreaming: "When I grow up, I'll be a rocker; I'll rock hard!"* (ERA, DK 40, 356)

Teachers may be surprised by children's knowledge of juicy expressions and imagery. The phraseologisms that children use in their direct sense, unaware of their generally known figurative meaning, can also be sources of jokes:

> *Marie (6): "Liisa did not come to pre-school today, because she is* ***warm in the head****"* (*'warm in the head'* in Estonian means stupid). (ERA, DK 40, 330)

The logic of acquiring the first and second language and the essence of the production process comes from the fact that a person's linguistic ability, the capacity to understand, attribute meaning and store knowledge is based on association and analogue-based generalisation, and it is directly related to the universal ability of self-organisation of the knowledge that is the basis of these processes (Eslon et al. 2010: 27). Many of the written recordings contain mix-ups of words, semantic changes or associations and analogies, which work as mnemotechnical methods in the case of children (children sort of remember something form earlier conversations, but not entirely correctly):

> *Samuel (5): "We went sleighing in* ***Jõgevapoeg hills****"* (he is referring to Kalevipoeg Hills near Jõgeva, merging them into Jõgevapoeg Hills). (ERA, DK 40, 202)

> *Teacher: "What river runs through Tartu?" Ott (6):* ***"Emavesi!"*** *"*(*Mother Water!"*) (the correct answer would be Emajõgi – Mother River). (ERA, DK 40, 202)

Repeated associations or linguistic developments that are probably influenced by pop culture (films and books) stand out among the material: for example, many children associate Dalmatian dogs with being black and white, calling black and white cows they see in pastures *Dalmatian cows* (ERA, DK 40, 12 (3.C)) and referring to a cabbage white butterfly as a *Dalmatian butterfly* (ERA, DK 40, 16); the TV show *Inspector Rex* has inspired them to use the name Rex as a synonym for a dog (ERA, DK 40, 11 (3.C)). There are two written records of *jäävhambad* (permanent teeth) being called *jäähambad* (ice teeth), which

is a typical example of semantic change (*jäädv* (permanent) is abstract, *jää* (ice) a concrete category) (ERA, DK 40, 244 (3.C) and ERA, DK 40, 327). This example is associated with the linguistic logic on the basis of which a child derives *leivahambad* (bread teeth) on the basis of *piimahambad* (milk teeth) (ERA, DK 40, 359); and it is also logical for a child that *neljapäev* (Thursday, literally the fourth day) is followed by *viiepäev* (fifth day; Friday in Estonian is actually *reede*) (ERA, DK 40, 358); etc.

Bilingualism has sometimes been the source of humour, as it is associated with semantic change and general knowledge, e.g. a child may see the word 'mum' and their mother's name as synonyms:

> *A girl from a Russian family is asked: "What's your mum called?"* ***"Mamotshka!"*** (Mum in Russian). (ERA, DK 40, 83)

Many texts illustrate the emphatic anthropomorphic approach of children whereby they liken the nature of surrounding objects to themselves, give life to lifeless objects, attribute feelings to lifeless objects, and so on.

> *Teacher: "What happens to a tree when rabbits chew off its bark?"*
> *Sigrid:* ***"The tree gets cold!"*** (ERA, DK 40, 356)

Children may also treat commodities, clothes, footwear, etc., as living beings in their speech.

> *Maria (4) comes in from outside: "I am all covered in snow and my* ***boots are cold****!"* (ERA, DK 40, 358)

In the Estonian version of the text, the girl uses the word *ümeli* instead of *üleni* when she says she is **all** covered in snow, which is humorous in itself, but the figurative image is a stronger method of making this sentence funny.

Many written records reflect the sincere attempts of children to justify and describe phenomena and situations on the basis of their experience and logic. Argumentation in these texts is extremely creative considering the limited life experience of children, but it may come across as extremely funny for adults:

> *Children are playing a guessing game. The teacher describes the object in the picture:*
> *"It's lifeless, round, and black and white."*
> *"A cow," guesses Alvaro.*
> *"I said that it's lifeless," says the teacher.*
> *"A dead cow!" says Alvaro.*
> (ERA, DK 40, 355)

Constant references to socio-cultural contexts

As we broaden the level of linguistic word formation, we see references to the child's socio-cultural experience in the remarks. Folklore emerges in the constant association with the surrounding cultural environment, so we can also say that the remarks made by children are a reflection of the surrounding world and they present genuine versions of this world, which have emerged on the basis of the child's life experience and the logic based on this. Eve Kikas (2010: 140) has pointed out that listening to the speech of adults and interpreting it is very important in the creation of new explanations; adults are the ones who supply words to children by answering their questions and helping them structure information. Unfortunately, the material is too random to make any ground-breaking conclusions about children's views of the world, but it does give an idea of how certain stereotypes are adopted at an early age. Therefore, the corpus contains meaningful reflections of gender stereotypes, the behavioural patterns children imagine a true man would exhibit and their understanding of how femininity is expressed.

> *Kalle tells the others with an air of importance: "A real man can take the cold!"* (ERA, DK 40, 354) (A real man can take anything.)

> *Raul to Madis-Mairold: "Guess why I don't like bald girls? Because I think then that they're boys!"* (ERA, DK 40, 354) (A girl who has short hair or is bald is boyish.)

> *Having heard a teacher say "Eww!" four-year-old Andri asks: "Are you also some shrieking woman?"* (ERA, DK 40, 330) (Women express themselves by shrieking.)

> *Five-year-old boy: "Women are necessary, because without them there would be no children."* (ERA, DK 40, 118 (3.C)) (The importance of a woman's role lies in giving birth and raising children.)

Many of the situations worth writing down arise spontaneously in the course of pre-school activities. The teacher's educational guidance can often be seen in dialogues, followed by the child's unexpected logic and interpretation of the situation:

> *Mattis (5) has put his hat on so that one flap is tucked in; the other one is hanging loose.*
> *Teacher: "A proper man doesn't go out like this."*
> *Mattis: "Yes he does if he wants to."* (ERA, DK 40, 11 (3.C))

In jokes, they place themselves in the roles of adults and try to imagine what they would do if they were grown-ups:

> *Karl (3): "I wish I was already grown up, so I could take myself a real wife."* (ERA, DK 40, 330)

The manner in which a child pictures and understands a situation has also been recorded in writing. The next story illustrates a child's genuinely emphatic impression of her father's place of work. The humour arises from the salience of a narrow detail that is incredible for adults (museum employees cannot go out to eat). The reasons why this impression was created remain 'behind the scenes', but would interest the researcher.

> *Haldi's father is the director of a museum. One day, after eating, Haldi takes two leftover biscuits from the table and says: "I'll take them to the ladies in the museum. They cannot go out to eat. I don't like starving people!"* (ERA, DK 40, 357)

The activities of adults, e.g. coffee drinking, also become noteworthy.

> *Four-year-old Piibe comes to pre-school in the morning and complains to the teacher, who is drinking coffee, that even though she ate at home, she's still hungry. When the teacher asks her if she would like some porridge, Piibe replies: "I don't want porridge, but I'd like some coffee!"* (ERA, DK 40, 333)

Children's sense of the world is best revealed in the repertoire that has been received in answer to the targeted questions and expectations of teachers. Teachers have specifically interviewed children and recorded the discussions of some topics, e.g. the whole group's discussion of the meaning of happiness and their opinions of what they would like to be when they grow up. The material is similar to interviews carried out by journalists about topical subjects (e.g. the article about celebrating Easter, written on the basis of an interview carried out in a pre-school (Päärt 2006)), and the targeted interviews carried out by researchers in order to clarify children's attitudes and perspectives regarding certain phenomena or their knowledge about a certain topic (e.g. Kikas 2010; Siim & Assmuth 2016). Technically, there are not as many linguistic jokes (although the linguistic base level may also be important), as the comedy arises from the difference between ordinary behaviour and the way a child understands the world on the basis of their limited life experience and how they express this understanding.

Interaction with the adult world is well illustrated by a discussion organised by a teacher, where children were asked who they wanted to be when they grow up.

I am enclosing the survey from 2010, where children were asked who they wanted to be when they grew up. These children were just about to leave pre-school for school.

> *Hester – I want to be a babysitter. I will show up at the door and ask: "Where's the baby?"*
> *Mattias – I will be a writer like my dad! You know, I changed my mind, I will be a general after all!*
> *Simoona – I can't really be bothered to become a doctor. I want to be a shop assistant!*
> *Anita – My mum wants me to be a doctor and I will! I have gentle hands!*
> *Jan-Jasper – I want to become a policeman.*
> *Kreete-Lisette – I don't know yet. Maybe a shop assistant... Pet shop assistant! I will change my mind.*
> *Loviisa Lorelai – I want to be a singer. If I cannot be a singer, then I'd like to be an accountant like my mum.*
> (ERA, DK 40, 226)

Humour arises in situations where the child imagines professions on the basis of their limited contacts with them, and emphasises certain details (e.g. you need gentle hands to be a doctor; or you can work in a pet shop when you become a shop assistant), and jumping suddenly from one thing to another (by naming professions that require cardinally different skills: writer/general, singer/accountant).

In some cases, issues that are very important for adults have been discussed. For example, the topic "How to save money in difficult times?", which was written down in Imavere pre-school in 2009, is clearly inspired by the recession that had hit society at the time:

> *Don't buy expensive things, like coffee machines.*
> *You can borrow from the bank.*
> *You only have to buy food.*
> *If you have a garden, you can get food from there.*
> *Don't go shopping at all.*
> *Just buy your things and then nothing else.*
> *My dad still has his job; he has money.*
> (ERA, DK 40, 361/2)

Complicated problems are discussed in pre-schools, and these discussions reflect children's knowledge of specific situations. This example reveals children's attitude towards money (you can borrow from the bank), consumption values

(children know that present-day coffee machines are expensive, food is of primary importance and it can be grown in a garden), and the topic of unemployment that is closely related to the social environment.

SUMMARY

The jokes written down by pre-school teachers represent the adult's position and, at the time of writing down, they were more adult than child humour, which made it important to use an interdisciplinary approach in the analysis, depending on the aspect. The way children experience and express things is usually turned into a story and written down by someone else, usually a family member – parent or grandparent, older brother or sister, child-minder, pre-school teacher, etc. – who happened to witness the event. Funny remarks meet all the conditions required to become folklore, group or family lore that binds a community, which may continue to spread verbally or in writing or become topical from time to time. Teachers record remarks with a sense of perspective, so the tradition forms an important verbal support to memories saved in baby photo albums and memory books.

The multi-layered records of children's remarks are closely related to humour and jokes in terms of their content, topics, and construction techniques, which is why the entire phenomenon can be analysed via theories of humour. According to theories of humour, remarks are usually accidental incidences of humour that arise in everyday activities, but sometimes they can also be the result of the targeted discussions of topics. The majority of remarks perceived as funny are based on slips of the tongue associated with the language acquisition process and linguistic development of children as well as the process of getting to know the world, which leads to unique and original interpretations. According to the GTVH, which was applied to the material, there are several levels that open and intertwine – the opposition of scripts, malfunctions in logical mechanisms, situationality and context, target of the remark, narrative strategy or genre, and linguistic resources. In accordance with the language acquisition process and linguistic development of children, the main emphasis in the analysis of children's remarks in this study is on the linguistic resource, including humour, where the collision between two opposing scripts is based on the lexical, including homonymic and patronymic, use of words and word formation. This is primarily expressed in pronunciation errors, mix-ups, and associative developments.

Children's remarks also give a rather direct reflection of the surrounding reality. The specific material does now allow for making profound absolute

conclusions about the child's view of the world, but it does make it possible to identify socio-cultural references, connections in the surrounding environment and constant (concealed) dialogue with the adult world (e.g. acquired cultural stereotypes, imitating the adult world), which can be studied further in the future with the help of additional interviews. It seems that the personality of the moment of recording is not important later on in the case of most jokes and the texts work as humour after the initial situation and context have dispersed.

Giving attention to what children say contests the well-known Estonian proverb, "A child speaks when a hen pees", and attributes importance to the child as a personality and his or her opinions. This approach is characteristic of the strong modern paradigm of childhood sociology, which focuses on the child. It can be said that such lore values little personalities and makes it possible to create emotional connections as the child grows (e.g. reduce the need for recognition during the teenage years), and in ideal cases could help different generations find a common language.

ACKNOWLEDGEMENTS

Research for this article was supported by the Centre of Excellence in Estonian Studies (CEES, European Regional Development Fund) and is related to research project IUT 22-5 (Estonian Research Council). The author wishes to thank the anonymous reviewers for the valuable comments.

NOTES

[1] The first collection of children's songs (58,832 pages) by Walter Anderson, the widely acclaimed folklorist and first professor of folklore at the University of Tartu, dates back to 1921–1939.

[2] Anu Vissel is the only folklorist in Estonia who regarded childlore as a separate field of research at doctoral level in her thesis "Children's Lore in the Changing Society" (Vissel 2004).

[3] The plans and materials of the training seminars are available online for officials and teachers engaged in pre-school education at http://folklore.ee/kp/lp/; materials about the nationwide competition in collecting pre-school lore held from 2010–2011 and the questionnaire can be found at http://folklore.ee/kp/lp/index_e.html.

[4] All of the material received during the collection competition has been entered in the database http://folklore.ee/Lapsesuu/ (Voolaid 2015), which is password-protected due to ethical reasons.

ARCHIVAL SOURCES

ERA, DK – Estonian Folklore Archives, collection of digital manuscripts

REFERENCES

Argus, Reili 2003. Lastekeelest ja selle uurimisest. [On the Child Language and Its Research.] *Oma Keel*, No 1, pp. 26–32. Available at http://www.emakeeleselts.ee/omakeel/2003_1/OK_2003-1_03.pdf, last accessed on May 12, 2016.

Argus, Reili 2007. Eesti lastekeelekorpuse morfoloogilisest märgendamisest. [Morphological Coding of Estonian Child Language Database.] In: Pille Eslon (ed.) *Tallinna Ülikooli keelekorpuste optimaalsus, töötlemine ja kasutamine*. Tallinn: Tallinna Ülikooli kirjastus, pp. 65–86.

Attardo, Salvatore 1994. *Linguistic Theories of Humor*. Berlin & New York: Mouton de Gruyter.

Attardo, Salvatore 2010 [2008]. A Primer for the Linguistics of Humor. In: Victor Raskin (ed.) *The Primer of Humor Research. Humor Research 8*. Berlin & New York: Mouton de Gruyter, pp. 101–155. DOI: 10.1515/9783110198492.101.

Attardo, Salvatore & Raskin, Victor 1991. Script Theory Revis(it)ed: Joke Similarity and Joke Representation Model. *HUMOR: International Journal of Humor Research*, Vol. 4, No. 3/4, pp. 293–347. http://dx.doi.org/10.1515/humr.1991.4.3-4.293.

EKSS 2009 = Langemets, Margit & Tiits, Mai & Valdre, Tiia (eds.) *Eesti keele seletav sõnaraamat.* [Explanatory Dictionary of Estonian.] Vol. 3, L–O, Tallinn: Institute of the Estonian Language. Available at http://www.eki.ee/dict/ekss/index.cgi?Q=lapsesuu&F=M, last accessed on May 12, 2016.

Eslon, Pille & Õim, Katre & Kaivapalu, Annekatrin & Argus, Reili & Matsak, Erika 2010. Kuidas uurida esimese ja teise keele omandamist? [Perspectives of the Research of the First and Second Language Acquisition.] *Lähivõrdlusi. Lähivertailuja*, Vol. 20, pp. 11–48. DOI:10.5128/LV20.01.

Kharchenko, Vera 2012. *Korpus detskikh vyskazyvanii*. [Corpus of Children's Remarks.] Moskva: Izdatel'stvo Literaturnogo Instituta im. A. M. Gor'kogo.

Hiiemäe, Mall 2014. Naljakatel eluseikadel rajanevad lood kogukonnapärimuses. [Narratives about Humorous Life Events in the Community Tradition.] *Keel ja Kirjandus*, No. 11, pp. 845–861. Available at http://kjk.eki.ee/ee/issues/2014/11/558, last accessed on May 12, 2016.

Kikas, Eve 2010. Laste mõtlemine. Pilved, vihm ja vikerkaar laste seletustes. [Children's Thinking: Clouds, Rain, and Rainbow in Children's Explanations.] *Mäetagused: Hüperajakiri*, Vol. 46, pp. 139–156. DOI:10.7592/MT2010.46.kikas.

Kraav, Inger 2007. Lapse areng valdkonniti ja võimalused sellele kaasa aidata. [Areas of Child Development and How to Help Them.] In: Siiri-Liisi Läänesaar (ed.) *Lapsehoidja käsiraamat*. Tallinn: Perekasvatuse Instituut, pp. 58–123.

Krikmann, Arvo 2004. *Koestler, Raskin, Attardo ja teised: Lingvistiliste huumoriteooriate uuemaist arenguist*. [Koestler, Raskin, Attardo and Others: New Developments of Linguistic Theories of Humour.] Reetor 4. Tartu: Eesti Kirjandusmuuseum.

Kõiva, Mare (ed.) 1995. *Lipitud-lapitud: Tänapäeva folkloorist 1*. [Contemporary Folklore 1.] Tartu: Eesti Kirjandusmuuseum & Eesti Keele Instituut.

Laineste, Liisi 2003. Suunad huumori uurimises. [Trends in Humour Research.] *Keel ja Kirjandus*, No. 11, pp. 793–804. Available at http://www.digar.ee/viewer/et/nlib-digar:195649/143197/page/3, last accessed on May 12, 2016.

Martin, Rod A. 2007. *The Psychology of Humor: An Integrative Approach*. Burlington: Elsevier Academic Press.

Martson, Ilona 2010. *Kükitav mannatera ja teisi eesti laste nalju*. [Squatting Manna Grain and Other Jokes of Estonian Children.] Tallinn: Tänapäev.

Mägiste, Julius 1924. Paar sõna lastekeelest. [Some Words about Child Language.] *Eesti Keel*, No. 1, pp. 1–9.

Nilsen, Alleen Pace & Nilsen, Don Lee Fred 2000. *Encyclopedia of 20th-Century American Humor*. Phoenix: Oryx Press.

Prout, Alan & James, Allison 2005. A New Paradigm for the Sociology of Childhood? Provenance, Promise and Problems. In: Chris Jenks (ed.) *Childhood: Critical Concepts in Sociology I.* London & New York: Routledge, pp. 56–80.

Põldmäe, Rudolf 1941. *Eesti rahvanaljandid I: Inimese eluperioodid*. [Funny Estonian Folk Tales I: Periods of Human Life.] Tartu: Teaduslik Kirjandus. Available at http://www.folklore.ee/rl/pubte/ee/nali/, last accessed on May 12, 2016.

Päärt, Villu 2006. Lihavõttejänes käib kodus mune peitmas. [Easter Bunny Is Hiding Eggs at Home.] Newspaper *Postimees*, April 15, p. 6. Available at http://www.postimees.ee/1539573/lihavottejanes-kaib-kodus-mune-peitmas, last accessed on May 12, 2016.

Raskin, Victor 1985. *Semantic Mechanisms of Humor*. Synthese Language Library 24. Dordrecht & Boston & Lancaster: D. Reidel Publishing Company.

Schmidt, Sigrid 2005. Kindermund – Erheiternde Kinderaussprüche und ihr Verhältnis zu Witzen. *Folk Narrative Theories and Contemporary Practices: Abstracts*. 14th Congress of the International Society for Folk Narrative Research (ISFNR). July 26–31, 2005, Tartu, Estonia. Tartu: Estonian Literary Museum & University of Tartu, p. 257. Available at http://www.folklore.ee/isfnr/abstracts.php, last accessed on May 12, 2016.

Siim, Pihla Maria & Assmuth, Laura (forthcoming). Mobility Patterns between Estonia and Finland: What about the Children? In: Kirsti Jõesalu & Anu Kannike (eds.) Cultural Patterns and Life Stories. Acta Universitatis Tallinnensis, Socialia / Humaniora. Tallinn: Tallinn University Press.

Chukovsky 2001 [1958] = Chukovskii, Kornei. *Sobranie sochinenii v 15 t. T. 2: Ot dvukh do piati.* [Collected Works in 15 Volumes. Vol. 2: From Two to Five.] Moskva: Teppa – Knizhnyi klub. Available at http://www.chukfamily.ru/Kornei/Prosa/Ot2do5/Ot2do5.htm, last accessed on May 12, 2016.

Vissel, Anu 2004. *Lastepärimus muutuvas ühiskonnas*. [Children's Lore in the Changing Society.] Ars Musicae Popularis 15. Tartu: Eesti Kirjandusmuuseumi etnomusikoloogia osakond. Available at http://dspace.ut.ee/handle/10062/1042, last accessed on May 12, 2016.

Voolaid, Piret 2012. *Meie armastuse emmed: Lapsesuu räägib*. [Our Beloved Mums: Child's Mouth Talks.] Tallinn: Ajakirjade Kirjastus.

Voolaid, Piret 2015. *Lapsesuu.* [Child's Mouth.] Tartu: Eesti Kirjandusmuuseum. Available at http://www.folklore.ee/Lapsesuu, last accessed on May 12, 2016 (password-protected due to ethical reasons).

http://dx.doi.org/10.7592/FEJF2016.64.velmezova

THE ESTONIAN LANGUAGE AS PRESENTED IN THE *GREAT SOVIET ENCYCLOPAEDIA*: LANGUAGE DESCRIPTIONS AND IDEOLOGY[1]

Ekaterina Velmezova

Abstract: Articles about the Estonian language published in the three editions of the *Great Soviet Encyclopaedia* bear the imprint of their epochs and ideologies. It becomes apparent even in the fragments of articles that contain language description 'as such' and should have been the most neutral ones. The article about Estonian from the first edition of the encyclopaedia seems the most ideologically marked: the view of Estonian in the general social context evidently prevails, while the language itself is only described to a lesser degree. This can be explained by political and social reasons.

Keywords: Estonian language, *Great Soviet Encyclopaedia*, linguistics and ideology, linguistic theories

At first sight, it may seem that articles about languages published in universal dictionaries and encyclopaedias should be extremely concise and impartial.[2] However, it turns out that these language descriptions are ideologically marked, to a greater or lesser extent;[3] for instance, they can reflect particular views of human language in general, which prevailed in the epochs and countries in which they appeared. This article demonstrates it on the example of articles about the Estonian language published in three editions of the *Great Soviet Encyclopaedia*.

The Great Soviet Encyclopaedia (hereinafter the GSE) earned itself the reputation of the most complete and well-known universal Soviet encyclopaedia. It was published in three editions: the first one (1926–1947) consisted of 66 volumes,[4] the second one (1949–1958) of 52 volumes,[5] and the third one (1969–1981) of 31 volumes.[6]

Articles titled "The Estonian Language" and published in these three editions (D. B. & K. B. 1934, *nomen nominandum* 1957b, and Ageeva 1978, correspondingly) differ from each other considerably, which could already be seen on the example of their authorship. In the GSE third edition the author's name is provided: the article was written by Ruf' Aleksandrovna Ageeva (born 1938);

in the second edition the article was published without any indication of its authorship under the text of the article; as to the first edition, the initials D. B. most probably refer to the name of Dmitrii Bubrikh (1890–1949), one of the most prominent Soviet scholars specialising in Finno-Ugric studies.

The volume of these three articles also varies from one text to another. The first one contains only 37 lines, the second (the most detailed one) – 99 lines, and the third, the most laconic one – 34 lines.

The articles are also very different as to their content and the manner in which they present the Estonian language: it is the article from the first edition that seems the most peculiar, considering all the parameters.[7]

FROM 'LINGUISTICS AS SUCH' (?) …

Estonian dialects and native speakers

In the article from the GSE first edition, Estonian is described irrespectively of its native speakers: "Estonian language (eesti keel) belongs to the Ugro-Finnic language system" (D. B. & K. B. 1934: 720) (about the notion of system in this context see below). On the contrary, in the GSE second and third editions, Estonian is defined, first of all, as a language of particular (native) speakers; here it is evidently presupposed that (the) language is first of all a means of communication. Besides, at the time when the second and third editions were published, Estonia was already a part of the USSR;[8] therefore it made sense to speak about the native speakers of Estonian in connection with the topic of the "Soviet population" in general.[9] In the GSE second edition, the Estonian language is defined simply as the language of Estonians (*nomen nominandum* 1957b: 225), while in the article from the third edition we find some more detailed information: it is the "language of Estonians living in the Estonian Soviet Socialist Republic, in the Leningrad, Pskov, and Omsk oblasts and in other regions of the Russian Soviet Federative Socialist Republic, in the Latvian and Ukrainian Soviet Socialist Republics, in the Abkhaz Autonomous Soviet Socialist Republic, and also in Sweden, USA, and Canada. The total number of speakers in the USSR is 1007,4 thousands [*sic*] of people (1970, [general] census)"(Ageeva 1978)[10] (information about the total number of Estonian speakers was also absent in the articles from the two first editions of the GSE; the question of the geographical spread of Estonian was not raised there either).

As to the main Estonian dialects, they are discussed in all three articles, but each time a particular dialectal division of Estonian is proposed. In the article from the first edition, two dialects [*narechie*] are mentioned: "The Es-

tonian language […] is divided into two dialects that differ considerably from one another: the northern, or the Revel one, and the southern, or the Dorpat one" (D. B. & K. B. 1934: 720). In the article from the GSE third edition, three dialects are discussed:

> *The Estonian language* […] *has three basic dialects: the north-eastern coastal, the north-Estonian and the south-Estonian ones. Beginning in the 16th century, there existed* [in Estonian] *two literary languages on the basis of the north-Estonian and the south-Estonian dialects*. (Ageeva 1978)

The article from the GSE second edition represents some kind of intermediary version between the articles from the first and the third editions. Like in the article from the third edition, three dialects [*narechie*] are mentioned here, but as the basic dialectal opposition in this article is also mentioned that of the 'north' and 'south', supposing the binary dialectal division of Estonian:

> *The Estonian language* […] *is divided into three dialects* [narechie]*: the northern, the north-eastern and the southern ones, which, in their turn, are divided into dialects* [dialekt] *and patois* [govor]. *The south-Estonian dialect, the basis of which was constituted by an ancient independent Baltic-Finnic language, is manifestly different from the two others, going back to another ancient language. The middle northern dialect* [srednii dialekt severnogo narechiia] *was chosen as the basis of the national literary Estonian language. Before the early 20th century, books were also printed in the south-Estonian dialect. At present, dialects tend to disappear.* (*Nomen nominandum* 1957b: 226)

Different interpretations are obviously given here to a similar situation, and this allows us once again to raise the question about the objectivity in language descriptions – even in those that seem the most neutral.

About the Estonian written language

From the GSE first edition we learn that "the Estonian written language (using Roman alphabet letters)", "because of the historical type of spelling, conveys the sounds of Estonian in an imperfect way" (D. B. & K. B. 1934: 721). In the third edition, there is no criticism at all (indeed, at that time Estonia was part of the USSR, so it was now a question about one of the Soviet peoples' written language): only the Roman letters of the Estonian alphabet are mentioned in this regard (Ageeva 1978). In the article from the GSE second edition, even the Estonian Roman letters are not discussed. Instead, some historical factors are

quoted: "The first records in Estonian go back to the 13th century (several sentences and glosses). More developed texts are known from the 16th century. The first book in Estonian was published in 1535" (*nomen nominandum* 1957b: 226).

On the history of the study of Estonian

In the GSE first edition, we find no information about the study of Estonian in the article "Estonian Language" as such; however, it is implicitly present in the references. In particular, the following sources are mentioned here: 1) Stahl[11] H[einrich], *Anführung zu der estnischen Sprache*, Reval 1637 "(the first grammar with a dictionary attached)"; 2) Wiedemann F[erdinand Johann], *Estnisch-deutsches Wörterbuch*, Petersburg 1869 "(still the most complete dictionary, re-edited in Saint-Petersburg in 1869 and in Tartu in 1923)"; 3) *id.*, *Grammatik der estnischen Sprache*, Petersburg 1875 "(the first scientific grammar)"; 4) Kuusik T[imoteus], *Vene-Eesti sõnaraamat* (Russian-Estonian dictionary), Tallinn 1906, and *Eesti-Vene sõnaraamat* (Estonian-Russian dictionary), Tallinn 1914; 5) Kettunen L[auri Einari], *Viron kielen äännehistorian pääpiirteet*, Helsinki 1917 "(the principal stages of the history of the Estonian language sounds; in Finnish)"; 6) Jõgever F. [*sic*] [Jaan][12], *Eesti keele grammatika* (Grammar of the Estonian Language), r.r. I–III, Tartu 1919–20. Finally, we learn that 7) the journal *Eesti keel* (Estonian Language) deals with the problems of Estonian and is published in Estonia since 1922 (D. B. & K. B. 1934: 721).

In the GSE second edition, in the references that accompany the article, in addition to the already mentioned (in the GSE first edition) books – F. Wiedemann's grammar and dictionary – we find two new sources by L. Kettunen (Kettunen L[auri Einari], *Vatjan kielen äännehistoria* (Phonetic history of Votian), [s.l.], 1929; *id.*, *Lauseliikmed eesti keeles* (Parts of the sentence in Estonian), Tallinn 1924), as well as some books published in the USSR – both in Russian (Piall' [Päll] E[duard], *Uchebnik estonskogo iazyka* (Course of Estonian), Tallinn 1955; Arumaa P[eeter], Pravdin B[oris], Veski I[Johannes] V[oldemar], *Russko-estonskii slovar'* (Russian-Estonian dictionary), Vol. 1–10, Tartu 1940–47; these books were the first in the Bibliography), and in Estonian (Kask A[rnold], *Võitlus vana ja uue kirjaviisi vahel* (Fight between new and old spelling), Tartu 1946; Ariste P[aul], *Eesti keele foneetika* (Phonetics of the Estonian language), Tallinn 1953) (ibid.).

In the article published in the GSE third edition, the only mention about the study of Estonian in its historical aspect is that "the Estonian language was unified [*unifitsirovan*] in the first half of the 20th century". Even the references to this article (unlike the bibliography to the articles published in the

GSE first and second editions) contain information reflecting, first of all, the study of Estonian at the time when this article was published. Here we find only four sources: the first two were published in Russian (Kask A[rnold], *Estonskii iazyk* (Estonian language), in the book: *Iazyki narodov SSSR* (Languages of the peoples of the USSR), Vol. 3, Moscow 1966, and *Osnovy finno-ugorskogo iazykoznaniia: Pribaltiisko-finskie, saamskii i mordovskie iazyki* (The foundations of Finno-Ugric linguistics: The Baltic-Finnic, Lappish and Mordvinian languages), Moscow 1975), and the last two – in Estonian, in the Estonian Soviet Socialist Republic (Ariste P[aul], *Eesti keele foneetika* (Phonetics of the Estonian language), 2nd ed., Tartu 1966, and Tamm J[ohan], *Eesti-vene sõnaraamat (*Estonian-Russian dictionary), Tallinn 1961 (Ageeva 1978). Foreign bibliographical sources are not provided in this bibliography at all, which implies mistrust in the quality of study of Estonian abroad – in contrast to its study in the USSR.

The language 'as such'

Nothing is said about the inner structure of Estonian in the article from the GSE first edition. On the contrary, the article from the second encyclopaedia edition contains the most complete (of all three GSE editions) description of the Estonian language 'as such'. As to the third edition, it seems to contain, in a contracted form, the information that has been developed in detail in the previous edition – with several exceptions, when the second and the third editions provided different linguistic data. This was the case, in particular, when the different numbers of consonant sounds in Estonian were indicated in these two articles[13] or when more attention was devoted to the particular Estonian prosody in the GSE second edition (*nomen nominandum* 1957b: 226).

As concerns the morphology, in the GSE third edition the Estonian language is described as "inflectional-agglutinative" [*flektivno-aggliutinativnyi*] (Ageeva 1978): in the 1970s, linguistic typology developed in the USSR in a very dynamic way, which is reflected in the use of terminology. Meanwhile, in the GSE second edition we still find no particular terms to refer to the language types; however, we learn that in Estonian "the name formally has only one type of declination and the verb has only one type of conjugation" (*nomen nominandum* 1957b: 226): it would presuppose a purely agglutinative structure of the Estonian language, which is not the case.

Speaking about the Estonian vocabulary, authors of all three articles emphasise the presence of loan-words in the language. However, in this regard, articles from the second and third editions are very similar and remain factual: "In the

Estonian language, like in all other Balto-Finnic languages, there are ancient loan-words borrowed from the Baltic, Finnic, and Slavic languages" (*nomen nominandum* 1957b: 226); "Estonian vocabulary includes ancient Finno-Ugric, Baltic-Finnic and properly Estonian words, together with many ancient Baltic, German, and Slavic loan-words (more recent loan-words are from German, Russian, and other languages)" (Ageeva 1978). As to the article from the GSE first edition, the presence of loan-words in Estonian is negatively connoted: in particular, we learn that "the Revel dialect" ("which is the basis of Estonian") "is more and more flooded with elements from the Finnish language" (D. B. & K. B. 1934: 721). This affirmation would seem impossible either in the second or in the third GSE editions: they described Estonian that was already a language of one of the Soviet socialist republics.

So, as we have already stated before, the GSE first edition contained very little information about the Estonian language 'as such'. In return, the content of this article manifestly reflects some linguistic theories that dominated in the USSR in the epoch when this text was composed and that were directly connected with a whole series of extralinguistic ideological parameters.

…TO IDEOLOGY REFLECTED IN LANGUAGE DESCRIPTIONS

On the 'system' *versus* 'family' of Finno-Ugric languages

In the GSE second and third editions, the Estonian language is defined as to its origins: it is "the language of Estonians, which belongs to the Finno-Ugric languages"[14] (*nomen nominandum* 1957b: 225); "the language of Estonians […]. It belongs to the Baltic-Finnic group of Finno-Ugric (Ugro-Finnic) languages" (Ageeva 1978). As to the article from the first edition, Estonian is defined here as a language that "belongs to the Ugro-Finnic language *system* [author's emphasis] […], more precisely, to its western-Finnic group" (D. B. & K. B. 1934: 720). This statement betrays the influence of N. Marr's (1864–1934) linguistic theories upon the authors of the article: in 1923–1924, Marr gave up the very idea of the existence of language families and the notion of language families itself, replacing it with that of a stage in the language development, or with that of a system.[15] Bubrikh, one of the supposed authors of this article, was known to be under a certain influence of Marr: his many researches can be regarded as a result of an attempt to find a compromise in order to reconcile Marrism with 'traditional' comparative linguistics. Unlike Marr, Bubrikh did not completely refute the genetic relationship between languages; however, under the strong pressure of Marrists he had to make many theoretical concessions.[16]

Besides, according to Bubrikh, in the constitution of languages and peoples (including the Finno-Ugric ones), the areal factors play a considerable role.[17] Likewise, in the article from the GSE first edition we find the thesis that "mass [*massovyi*] and extended contacts with [other] peoples – Germans, Russians, and Swedes – left deep traces in the [Estonian] language, changing not only its vocabulary, but also its morphology and phonetics. The most evident traces are those of the Baltic barons' domination" (D. B. & K. B. 1934: 721). The last statement allows to proceed to another topic that is manifestly present in this article – the notion of class nature of a language or class distinctions in any (natural) language. This question was actively discussed in the Soviet linguistics of the 1920s.

Estonian 'barons' language' and Estonian language of 'working and peasant population'

In particular, the above-mentioned Baltic barons' domination and its influence on the Estonian language was supposedly connected with the fact that "with the purpose of influence over the peasant masses, it is the German pastors who had elaborated the Estonian written language. The first manuscripts in Estonian go back to the middle of the 16th century, but only in the second half of the 19th century, during the development of the literary Estonian language there was a turn towards the regulation [*uporiadochenie*] and towards the enrichment [*obogashchenie*], in connection with the fact that literature became an instrument of the Estonian national bourgeoisie that emerged at that time. A new turning-point in the development of the Estonian literary language can be connected with the emergence of the Estonian bourgeois republic, when literature became an instrument of the dominating class. i.e. the bourgeoisie. Now the Estonian language is elaborated without any connection with the demands of working and peasant masses; for a moment, the northern (Revel) dialect has been finally [*sic*] established as its basis, at the same time it is more and more flooded with the Finnish language elements. In connection with this, in the Estonian circles of the USSR, the idea about the creation of an Estonian literary language, which would be close to the language of working and peasant masses, grows stronger and stronger" (D. B. & K. B. 1934: 721).

In the Soviet Union of the 1920s–1930s, the idea of class distinctions in the language runs all through the works of both obvious Marrists and linguists who did not unambiguously belong to the Marrist paradigm; sometimes even the existence of national languages as such was denied (for instance, in G. K. Danilov's works (1929a, 1929b)). From time to time, Marr himself also refuted the exist-

ence of 'national languages' (cf. Marr 1933–1937, II: 415), maintaining that only 'class languages' exist and that, for instance, languages of representatives of the same class in various countries are more similar than the languages of different classes of the same people (ibid.). However such statements did not prevent Marr from writing about the Estonian, Russian, and other 'national languages', which shows to what extent Marr's theories were contradictory. The division of language into 'class languages', without the denial of the very idea of the 'national language', is *de facto* also presupposed in the article about Estonian published in the GSE first edition.

CONCLUSION

As we can see, even the seemingly neutral articles about the Estonian language published in the GSE are ideologically charged, reflecting the particular political and social situation of the epochs when they were composed. It can be seen both in the references accompanying the articles (Soviet of foreign sources quoted) and in particular fragments of the language description (are loanwords or the written language as such criticised, or the attitude towards them is neutral, etc.). In this regard, the article from the GSE first edition is the least neutral: it reflects the influence of Marrism on its authors; the language 'as such' is described very little, while the extralinguistic situation around Estonian is presented rather negatively. It can be explained by the fact that, in the early 1930s, Estonia was not yet a part of the USSR, which made it possible to criticise the 'non-Soviet' Estonian language, its dependence on the dominating class of the bourgeoisie, the flood of Finnish elements in Estonian, etc. In addition to a particular (scientific) ideology, all this reflects the style and language of the Soviet linguistics itself during the same epoch.

ACKNOWLEDGEMENT

I would like to thank Prof. Kalevi Kull for his helpful and critical comments.

NOTES

[1] To Prof. Urmas Sutrop on the occasion of his 60th birthday, with much gratitude for his work.

[2] The question about the 'impartiality' in language descriptions presents a particular problem first of all in connection with the opposition 'ontological *versus* constructed (object of study)'.

[3] By *ideology* we mean here an aggregate of doctrines, opinions or beliefs shared by members of a group.

[4] In the GSE first edition, there were 65 volumes and the additional volume titled *The USSR*, which had no number.

[5] The GSE second edition contained 49 volumes, the 50th volume titled *The USSR*, the additional 51st volume and the volume 52, *Alphabetical Index*, edited in two books (1960).

[6] One of the thirty volumes of the GSE third edition was published in two books (the additional book under the title *The USSR*), to which was added the volume *Alphabetical Index of Names* (which had no number).

[7] Taking into account the limited volume of this article, I am not going to discuss *all* fragments of the Estonian language descriptions presented in the GSE, but only those that seem particularly important within the framework of my general subject, "Linguistics and ideology".

[8] Cf. in this regard the following quotation: "A new epoch [*etap*] for the Estonians began after the restoration of the Soviet power in 1940. In the process of socialist construction, in the political, economic, and cultural collaboration with other peoples of the USSR, the Estonian nation has transformed into a socialist one" (Viires 1978).

[9] For instance, in the article titled "Estonians" and published in the GSE third edition, we read the following: "Estonians, a nation (in the second edition, 'socialist nation' (*nomen nominandum* 1957a)), basic population of the Estonian Soviet Socialist Republic" (Viires 1978). Cf. in this connection the definition of the Estonian language from the Concise Literary Encyclopaedia, also published in the USSR: "Estonian language is the language of Estonians, the basic population of the Estonian Soviet Socialist Republic" (Kask & Ageeva 1975: 978).

[10] Here the region of spread of Estonian coincides exactly with the one where the Estonians live, according to the article "Estonians" published in the same GSE edition: "Estonia, Russian Soviet Federative Socialist Republic (Leningrad; Leningrad, Pskov, and Omsk oblasts, etc.), Latvian Soviet Socialist Republic, Ukrainian Soviet Socialist Republic, Abkhaz Autonomous Soviet Socialist Republic, USA, Sweden, Canada" (Viires 1978) (in this way the sign of equality between *language* and the corresponding *people* was implicitly presupposed).

[11] Hereinafter I have added full names to the GSE spelling. The spelling of titles of all quoted sources corresponds to the GSE one. Source titles are italicised by the editor; additional bibliographical data and translations of the titles are also sometimes provided by the editor.

[12] Wrong first name in the GSE (*E.V.*).

[13] Precisely 'sounds', and not 'phonemes' were under discussion in these articles, even if at present the word phoneme would seem more appropriate in this context.

[14] Besides, unlike the GSE first and third editions, the second edition mentions the "nearest relations" of Estonian: they are Livonian and Votic languages (*nomen nominandum* 1957b: 225).

[15] With particular regard to the Estonian language, I analyse Marr's theories in the article *Estonskii iazyk glazami sozdatelia "novogo ucheniia o iazyke"* (Vel'mezova forthcoming).

[16] Cf., for example, in his *Autobiography*: "[…] with my bourgeois education, I could not immediately get free from the methodology of bourgeois linguistics in order to appreciate properly academician N. Marr's Japhetic theory [in this particular context, Japhetic theory is the same as Marrism] developing towards the Marxist-Leninist linguistics. Having accepted a series of Japhetic theory theses (for instance, those against the idea of protolanguages), I had harshly argued against others, for instance, against those that were based on the idea of the discovery of sound speech primary elements. Nevertheless, I should say that already by 1932, I had found the right line" (Bubrikh 1934). And the following is the description of Bubrikh's situation in the early 1930s by V. M. Alpatov: "[…] Bubrikh has also surrendered. And N. Troubetzkoy, who had once found in one of his works the 'signs of genius', was disappointed in Bubrikh's work about the Erzya-Mordvinian language [(…) 1935 (…)], where he refuted the principles of comparative linguistics" (Alpatov 1991: 93, cf. also 118, 128–129, 140, 199). Despite his many attempts to find theoretical compromises with Marrists, in spite of the fact that in 1934–1949 Bubrikh was in charge of the Finno-Ugric department at the N. Ia. Marr Institute of Language and Thought, he suffered much because of the Marrists – both because of their criticism in the early 1930s (ibid.: 101) and because of their persecutions in the late 1940s, which finally cost Bubrikh his life (ibid.: 158–159).

[17] Later it was emphasised, in particular, in the article of Bubrikh's disciple Georgii Martynovich Kert (1923–2009) on the example of Bubrikh's study of the Karelians' origin (Kert 2002; cf. also Alpatov 1991: 133, etc.).

REFERENCES

Ageeva, Ruf' 1978. Estonskii iazyk. [The Estonian Language.] In: Aleksandr Prokhorov (chief ed.) *Bol'shaia sovetskaia entsiklopediia*. 3rd edition, Vol. 30. Moscow: Sovetskaia entsiklopediia, p. 284.

Alpatov, Vladimir 1991. *Istoriia odnogo mifa: Marr i marrizm.* [The History of a Myth: Marr and Marrism.] Moscow: Nauka, Glavnaia redaktsiia vostochnoi literatury.

Bubrikh, Dmitrii 1934. *Zhizneopisanie Dmitriia Vladimirovicha Bubrikha*. [Dmitrii Vladimirovich Bubrikh's Biography.] Available at http://iling.spb.ru/pdf/liudi/bubrih.html, last accessed on May 17, 2016.

D. B. & K. B. 1934. Estonskii iazyk. [The Estonian Language.] In: Otto Shmidt (chief ed.) *Bol'shaia sovetskaia entsiklopediia*. 1st edition, Vol. 64. Moscow: OGIZ RSFSR, Gosudarstvennoe slovarno-entsiklopedicheskoe izdatel'stvo Sovetskaia entsiklopediia, pp. 720–721.

Danilov, Georgii 1929a. Iazyk obshchestvennogo klassa. [The Language of a Social Class.] *Uchenye zapiski Instituta iazyka i literatury RANION*, Vol. 3, pp. 163–194. Available at http://crecleco.seriot.ch/textes/DANILOV29/txt.html, last accessed on May 17, 2016.

Danilov, Georgii 1929b. Marksistskii metod v leksikologii. [The Marxist Method in Lexicology.] *Russkii iazyk v sovetskoi shkole*, No. 6, pp. 48–62. Available at http://crecleco.seriot.ch/textes/Danilov29b.html, last accessed on May 17, 2016.

Kask, Arnol'd & Ageeva, Ruf' 1975. Estonskii iazyk. [The Estonian Language.] In: Aleksei Surkov (chief ed.) *Kratkaia literaturnaia entsiklopediia*, Vol. 8. Moscow: Izdatel'stvo Sovetskaia entsiklopediia, pp. 978–979. Available at http://feb-web.ru/feb/kle/kle-abc/default.asp, last accessed on May 17, 2016.

Kert, Georgii 2002. Velichie i tragediia talanta. [The Grandeur and Tragedy of a Talent.] In: G. Kert *Ocherki po karel'skomu iazyku: Issledovaniia i razmyshleniia.* 2nd edition. Petrozavodsk: Kareliia, pp. 83–107. Available at http://www.voinitsa.ru/pages/art243.aspx, last accessed on May 17, 2016.

Marr, Nikolai 1933–1937. *Izbrannye raboty.* [Selected Works.] Vol. I–V. Moscow & Leningrad: Izdatel'stvo gosudarstvennoi akademii istorii material'noi kul'tury (Vol. I), Gosudarstvennoe sotsial'no-ekonomicheskoe izdatel'stvo (Vol. II–V).

Nomen nominandum 1957a. Estontsy. [The Estonians.] In: Boris Vvedenskii (chief ed.) *Bol'shaia sovetskaia entsiklopediia*. 2nd edition, Vol. 49. Moscow: Gosudarstvennoe nauchnoe izdatel'stvo Bol'shaia sovetskaia entsiklopediia, pp. 226–228.

Nomen nominandum 1957b. Estonskii iazyk. [The Estonian Language.] In: Boris Vvedenskii (chief ed.) *Bol'shaia sovetskaia entsiklopediia*. 2nd edition, Vol. 49. Gosudarstvennoe nauchnoe izdatel'stvo Bol'shaia sovetskaia entsiklopediia, pp. 225–226.

Vel'mezova, Ekaterina forthcoming. Estonskii iazyk glazami sozdatelia "novogo ucheniia o iazyke". [The Estonian Language as Viewed by the Creator of the "New Theory of Language".] In: Anastasiia de Lia Fortel' & Sebast'ian More (chief eds.) *Sbornik statei k 65-letiiu Patrika Serio.* Moscow: OGI.

Viires, Ants 1978. Estontsy. [The Estonians.] In: Aleksandr Prokhorov (chief ed.) *Bol'shaia sovetskaia entsiklopediia*. 3rd edition, Vol. 30. Moscow: Sovetskaia entsiklopediia, p. 284.

http://dx.doi.org/10.7592/FEJF2016.64.giants

GIANTS IN TRANSMEDIA

Mare Kõiva, Andres Kuperjanov

Abstract: The purpose of this article is to discuss transmedia narratives based on giant lore, which is described by means of examples from folkloristics and transmedia dissemination. Giant lore, particularly the epic *Kalevipoeg*, a core text of Estonian culture, has generated numerous transmedially circulating texts and various contemporary forms. Through their connections with media, texts about giants continue to participate in the national cultural space; in previous eras, they have been carriers of Estonian identity or, alternatively, have held an important place in the creation of local identities. The latter can be observed today in printed matter, advertisements, and products marketed to the homeland public. However, texts about giants can also be used as a self-characterising image directed beyond national space. The article provides a closer look at ways in which stories connected with Kalevipoeg and Suur Tõll are engaged in different levels of media, as well as necessary contextual cultural knowledge for understanding contemporary media clips.

Keywords: cross-media, giant lore, transmedia narrative

Media narratives scholar Marie Laure Ryan emphasises that some transmedia narratives are culturally representative or valuable stories which, when they acquire sudden or accelerated popularity, spontaneously attract new stories in a kind of snowball effect. Such stories spawn the creativity of fans and generate transmedia adaptations. In any event, such a media bundle has a central or core text that acts as a common reference for the remainder of the texts (Ryan 2008).

Without a doubt, one of the core texts of Estonian cultural space is the national epic *Kalevipoeg* (Kalev's Son), which over the past few centuries has performed a range of different functions for Estonians, including impelling them to argue about serious philosophical topics. Nevertheless, *Kalevipoeg* has preserved the outlines of a simple, explanatory folkloric text, which at different periods of time has been denounced and criticised, adapted or parodied, or reduced into prose retellings. It has inspired speeches in traditional verse form (*regivärss*) performed on festive occasions; likewise, it has generated amateur theatrical productions and professional elite performances. To put it another way, *Kalevipoeg* has provided an impulse for a myriad of different texts, corpuses of texts, adaptations, parodies, and rearticulations in different art forms which

have extended beyond the various media into the social space. Such flexible traffic between and among media is the outcome of a longer period during which the support of the Estonian community and the state in its different forms has played important roles.

How can we describe the spontaneous circulation of *Kalevipoeg* among various media and today's new emergent forms (computer games, memes, Kalevipoeg comics, videoclips, advertisement clips, etc.), a dynamic that is also manifested in hiking paths, dolls, and other products? Is it possible to treat the text of a canonical epic in terms of a structure of franchises? What was the nature of the connection between giant lore (on which the creation of the epic is based) and the power of text collections to circulate among various media forms, actualising themselves in society as phenomena similar to franchises?

In our brief characterisation of these phenomena, we have also made use of the term *transmedia narrative*, which was adopted by new media scholar Henry Jenkins. Using such a term in the present context is debatable, because up till now it has primarily been used to describe the most recent of contemporary texts, those with blockbuster potential. Yet at the moment we do not have a better alternative for characterising extramedial cultural and commercial phenomena.

Jenkins' approach was based on the observation that popular stories such as the Harry Potter series, *Matrix*, *Lord of the Rings*, and the like are able to cross the boundaries of a medium and assume new forms in other media. According to Jenkins' definition, this phenomenon entails storytelling on multiple platforms and in multiple formats, in other words, the flow of content through different platforms (Jenkins 2006: 2). In each channel of communication, a transmedia product develops a unique content, though the stories are linked among themselves and synchronised through their characters. Besides the creation of official versions by professionals, there are independently created variations, reworkings, adaptations, parodies, and other elaborations spread in social media environments. Jenkins was also one of the first to call attention to the ability of characters in a narrative to move beyond the boundaries of media, circulating transmedially in entertainment and commerce, for example, as franchises (branded products); as such these characters can play a variety of different roles in cultural space. The term *franchise* refers to theme parks, restaurants, airplanes (as well as other modes of transportation), clothing, and food brands – an innumerable wealth of products and services.

In his subsequent publications Jenkins has specified the main outlines of transmedia narratives. In his blog post "Transmedia Storytelling 101" (Jenkins 2007), he notes that what counts is the entertainment experience, adding that in the ideal case, the presentation in each medium makes a unique contribution. For example, the key aspects of information concerning the roles of the

characters of *Matrix* have been represented in three action films, short animated film series, two collections of comics, and several video games. No source – not even the original text – can represent or contain all the information required to understand the *Matrix* universe. Thus these are examples of transmedia projects in the contemporary market economy; alongside and in addition to these, volunteers, fans, and others create stories with the same characters and motifs (Ryan 2008; Jenkins 2004, 2007).

The terms transmedia and cross-media have been defined in, and distinctions made between them by new media theorists are often relative and labile. For example, for Moloney (2011, 2011–2014) the primary difference is that the term *cross-media* designates the distribution and creation of content (e.g. music, text, pictures, video, etc.) among different media, as well as an orientation toward economic success. Practitioner and visionary Günther Sonnenfeld (2009) regards reciprocal movements and influences as important facets of transmedia; while transmedia reflects activities and exchanges, cross-media tends to be more linear.

Assuming that transmedia is more connected with spontaneous cultural processes (the natural dissemination of a motif, character, etc. in culture), the term *transmedia* has seemed most appropriate to the purposes of the current study.

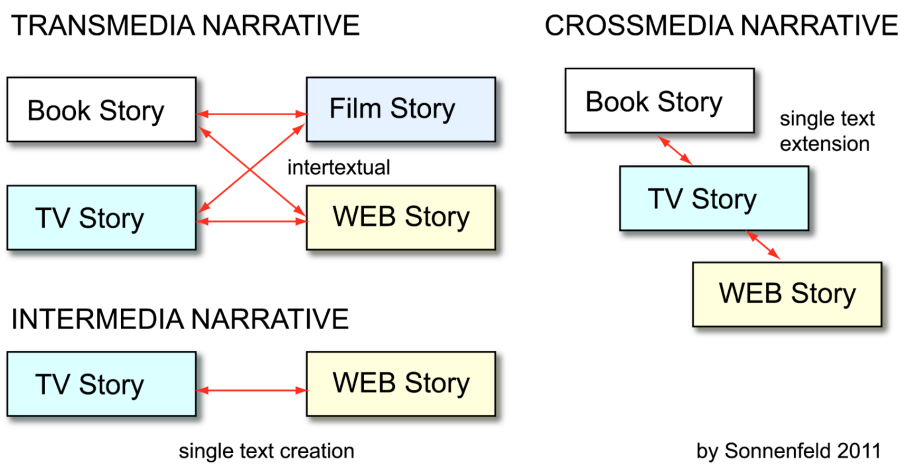

Figure 1. *Media narratives by G. Sonnenfeld 2011.*

FOLKLORE AND MEDIA

The basis of the storyworld of giant tales and the universe of giants consists of oral folkloric tales, in which heterogeneous folkloric characters perform various deeds and actions. Though the text aggregates around Kalevipoeg, Tõll, and other giants probably originated in or around the same time, there can be no comparison with respect to the number of texts, the abundance of motifs, or the geographical scope of the lore concerned. Kalevipoeg is primarily aligned with eastern Estonia; the Pagan (Old Nick) is known throughout Estonia; Tõll is known mostly in Saaremaa, and Leiger in Hiiumaa. The smaller the geographical area associated with the lore of a particular giant, the fewer themes about them have been preserved in the text aggregates.

The folklore collections are rich, but they do not cohere on the basis of content. Rather, old mythological motifs can be found in the collections, such as dualistic creation myths, in which the world is shaped as the outcome of a competition between the god and the devil (*pagan*). The origins of specific landscape formations are also linked to similar competitions, or regarded as traces of the activities of giants.

Giants have some features in common with primal beings of nature: they turn to stone at sunrise; they are huge in size; sometimes they are cannibals; they can be one-footed or one-eyed, like the Cyclops. They also have motifs in common with traditions about saints and other discourses of folk tradition. The majority of giant traditions focus on the giants' social organisation, and explanatory tales justify the origin of specific land forms and noteworthy natural objects. For example, in Estonia, which is full of stones, it is important to explain how the stones originated and how they came to be located where they are in the landscape.

Models for explaining this phenomenon often include Kalevipoeg's joy in throwing rocks and legends of the Old Pagans (*vanapaganad*) as carriers of stones. However, in western Estonia and the islands, rocks are carried by the wife of the giant (*vanapagan* or Tõll).

Folklore about giants made its way into print beginning in the 18th century by means of overview accounts of local life and customs written by historically-minded Baltic Germans. For example, the first lengthy account of Tõll, the giant of Saaremaa, was written by the Baltic-German linguist, pastor and intellectual August Wilhelm Hupel (1737–1819) in part III of his most important work, *Topographische Nachrichten von Lief- und Ehstland*. This work, published in 1782, continues to be cited frequently in scholarly publications. By contrast, the first reports concerning Leiger appear at the beginning of the 19th century (Luce 1827).

At the beginning of the 20th century, folklore texts, tales of origin, and explanatory tales began to appear in the Estonian language through the mediation of folklorists, such as the books of Matthias Johann Eisen and his series titled *Estonian Mythology* (Eisen 1901, 1910, 1913, 1920, 1924, 1926a, 1926b, 1926c, 1927, 1930, etc.). The two decades following the Second World War saw the publication of academic editions documenting the areas of dissemination of giant tales and describing the richness of their thematic range (Kalevipoeg stories by Laugaste & Normann 1959; about Tõll, Leiger, and others by Laugaste & Liiv & Normann 1963; about Vanapagan by Laugaste & Liiv 1970). Despite the fact that in some of these works, there is a tendency on the part of the compiler to connect smaller themes into a larger whole, these occurrences can be regarded as exceptions, as contaminations of the corpus by narratively unelaborated stories joined together mechanically.

One area of application of giant folktales relates to natural and archaeological monuments, where the exemplary folktale was incorporated into the legend of the monument. Thanks to monuments, folktales about giants came to have their place in nonfictional works on local history and thematic photographic albums, such as Helmut Joonuks' *Kalevipoeg's Places* (Joonuks 1982). Among printed texts intended for a smaller, interested readership, there are texts that pursue historical explanations for folktales and natural objects. One noteworthy author in this category is the Estonian language teacher Eduard Leppik, who uses drawings and comparisons with local history as an explanatory model, arriving at the conclusion that the landscape form spoken of as Kalevipoeg's horse actually refers to trenches dating back to the Russian-Swedish war (e.g. Leppik 2001). The value of such publications is the recreation of cultural history for mass consumption, and the creation of multiple meaningful connections between local landscape, folklore, and cultural history. Later on, creation stories have found their way into collections of local folktales; such volumes began to be composed at the end of the 20th century, and have continued to expand in the 21st century. Tales of giants' deeds can also be found in these works, as, for example, in the Mulgi Cultural Institute collection compiled by Kalle Gaston (1999) and authorised amplifications authored by Henno Käo (2001).

THE TRANSFER OF TÕLL INTO MEDIA

The 19th and 20th centuries in Estonia were significantly shaped by epic discourse. Bruce Lincoln (2000), a philosopher and researcher of myths, has called attention to a long-term cultural process during which the epic mode and the existence of epics became the measure of dignity for history and culture, along-

side the dichotomy of high and low. Besides the biblical lands and the cultural space of Greece and Rome, the 18th century translation of the *Poetic (Old Norse) Edda* valorised the culture of Germanic Scandinavia, creating a place for it in the 'circulatory system' of ancient high cultures. For Estonia, the core text – also a hit text – was the epic published by Fr. R. Kreutzwald in the years 1857–1861, based on narratives about giants and articulated in the form of alliterative folk poetry (*regivärss*). This epic was *Kalevipoeg*, extended and supplemented by themes of the author's own creation. To date, this is the most frequently translated work of Estonian literature into foreign languages,[1] resulting in the inclusion of Estonians among 'peoples of the epic'. At the time of its publication, *Kalevipoeg* was the longest original secular work of literature in the Estonian language. Thus the epic had an influence on general processes of self-consciousness and self-perception. It is reported to have been read out loud in the taverns, just like newspapers (Tedre 2003). One should not underestimate the significance of the fact that in the 19th century, excerpts from the epic were included in readers and Estonian-language textbooks; this was a crucial facet of the epic's transformation into a core text. In later periods, the epic has continued to be included in the school curriculum and it has become a focal point of texts generated by cultural professionals.

It must be remembered that a core text is also a canonical text. This means that it is possible to set oneself into opposition with it or become accustomed to it; one can elaborate on it and research it. For example, the folklore bibliography points out that the epic has been one of the most productive topics of research in the humanities (cf. Ribenis 2002).

Almost a century after the publication of the first Tõll stories, the poet, musicologist, and amateur scholar Martin Körber (1817–1893) wrote a German-language overview of them. Körber was known as a proponent of the theory of sung epics, which had been a major topic for discussion for almost a century. Tõll fulfilled Körber's expectations regarding the requirements for a national hero: he was a leader and elder for his people, a man of formidable strength, a hero and a warlord. Körber asserts as much in the second volume of his work, *Oesel Einst und jetzt* (Saaremaa Before and Now) (Körber 1887: 150ff.). Tõll's activities as a leader of his people are foreign to the folktale form, belonging rather to the writer's world of fantasy. The wish to confirm the lasting impact of the Tõll epic impelled Martin Körber to send Jakob Hurt repeated enthusiastic reports about the singers of the *Kalevipoeg* cycle and singers representing other longer epic cycles (cf. Laugaste & Liiv 1970).

Tõll's features as a leader of the people are foregrounded in the book, *A Little Box of Old Treasure, or Saaremaa's Hero, the Great Tõll,* written by Peeter Süda (1830–1893), a man of letters from Saaremaa. Süda was enthusiastic about the

opportunity to bring the heroic saga about the giant and defender of the people of Saaremaa to a broader public, using folkloric texts to do so. His ambition was to provide an appropriate local alternative to *Kalevipoeg*. This book was very popular among readers both on the islands and on the mainland, since the author had arranged the individual stories into a coherent whole; the book was easy to read, and it was important locally for construcing an image of 'our own hero'. Thus, albeit as a prose narrative, and almost thirty years after the publication of *Kalevipoeg*, the Tõll epic was born.

One of the cultural highlights of the 1980s was the screening of Rein Raamat's animated cartoon film *Suur Tõll* (Tõll the Giant). The scenario is based on Peeter Süda's Tõll book (Süda 1883) with additional ideas drawn from folklore collections. The images of Tõll and other characters were drawn for the film by the artist Jüri Arrak. A particularly psychedelic nuance was added by Lepo Sumera's soundtrack, the slow, epic quality of the action, and a very laconic script. A book based on the film, titled *Suur Tõll*, was published two years later with illustrations by Jüri Arrak (Raamat 1982). The two- or three-eyed devil and the clumsy Tõll provoked both enthusiasm and criticism, as well as a great deal of discussion. *Suur Tõll* was popular among the artistic elite, but the artist's representation of the *Suur Tõll* story was mesmerising as well as emancipatory, awakening young people's interest in folklore.

A critic holding the viewpoint of teachers and parents raised the question of whether it was right to frighten children, and whether it was ethical to show principal characters as nauseatingly revolting creatures (Niineste 2010). Arrak's pictures were dominated by blue-violet colour tones and the enemies were depicted as an anonymous mass wearing hats that belonged either to the Ku Klux Klan or lepers. As a whole, the visual representation successfully distanced itself from the romantic style of depicting ancient history, thus making room for new paths of interpretation.

While Peeter Süda cut a path for the folkloric giant to enter print media, Arrak's book had a major influence on its later versions. In 2014 popular writer Andrus Kivirähk published another Tõll book consisting of free-form texts narrated in a conversational tone. The author used imagery similarly to oral texts, with frequent allusions to the present day, telling the old stories in a contemporary style, with a generous helping of humour.

Kivirähk's book drew upon Jüri Arrak's cult illustrations. Arrak himself regards Kivirähk's text as having restored the power of nature to Suur Tõll: "At first he was a large horrible creature, but this text is more mystical, more powerful with regard to nature" (Ringvaade 22.10.2014). The writer ventured the modest claim that his was a book neither for children nor for grownups, best situated in the same category as Greek myths (ibid.).

GIANTS IN CONTEMPORARY MEDIA

Next, let us look more closely at a music clip and two advertising clips characterised by a plentiful use of cultural quotations. The first is an advertisement for consumer protection; the second for a series of Kalev[2] chocolates (the Epic series).

In her analysis of visuality as a part of phraseology and advertisements, Anneli Baran has emphasised the importance of multiple meanings in the perception of the text. These meanings are deployed on three levels: a) the level of quasi-visual imagination located in the individual; b) the level of knowledge (frames, scenarios, scripts); and c) the conceptual-metaphorical level (Baran 2008). In the present case, allusions are created by the pictures, though the level of knowledge, that is, cultural and semantic literacy, is also important. Anneli Baran points out that when interpreting an unknown expression, the viewer relies on associations, and the resultant explanation is more or less aligned with the most contemporary meaning (ibid.).

A singer in the Estonian ethno-rock band Metsatöll has confessed to a deep fondness for Arrak and Raamat's animated cartoon film, which influenced him in his youth.

Metsatöll's music clip *Only Bravery* (2010)[3] is built up around cultural quotations: one text is used to amplify the message of another. In the background of the musical piece we see the musicians performing in alternation with excerpts from Rein Raamat and Jüri Arrak's animated cartoon film. The visual background supports the band's rugged music; two well-known symbols are interwoven, resulting in a synthesis.

By contrast, in an advertising clip sometimes many verbal and visual styles can be mixed together. This can be seen in the consumer protection advertisement, *Kalevipoeg Performing for the Metsatöll band*[4]. In the text, imitations and improvisations of folk poetry (*regivärss*) alternate with ordinary language, while the visuals narrate the meeting between the epic character Kalevipoeg and the cult band – Kalevipoeg's failed trial performance for the rock musicians.

The blond, long-haired Kalevipoeg, dressed in a stylised ancient robe, plays an electric *kannel* (zither). Suddenly, the strings break, accompanied by abundant puffs of smoke, and just as in the epic, the slight hero receives advice from a hedgehog. Instead of striking the enemy edgewise with boards, the viewer hears a reminder that Consumer Protection Board can help in a bad situation.

The frames of the slight Kalevipoeg, the electric *kannel* and the hedgehog alternate with the Metsatöll singers, depicted as serious-faced giants dressed in black. They are all listening to and evaluating Kalevipoeg's musical performance. On the visual level, self-referential and self-generative characters

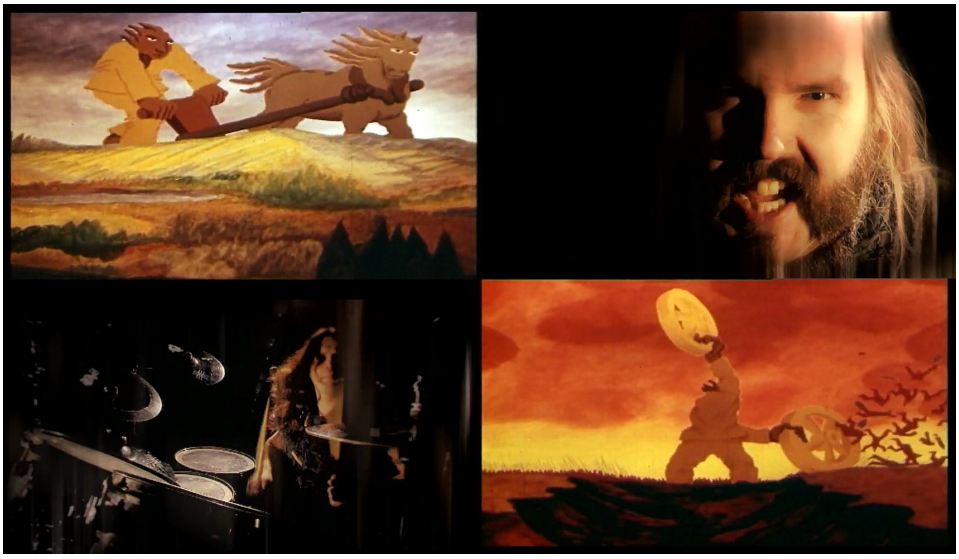

Figure 3. *Metsatöll,* Vaid Vaprust *(Only Bravery) on Youtube, using images from Suur Tõll.*

Figure 4. *Images of Kalevipoeg. 1–2. Kalevipoeg is doing audition for the Metsatöll band. Advertisement for Consumer Protection Board. 3–4. Tastes to remember (Kalev chocolate advertisement).*

are placed side by side: the cult band, the weak Kalevipoeg, the practical, matter-of-fact hedgehog. The fact that all the components and characters are generally known supports the multiplication of allusions, comic elements, and shared interpretations.

The text of the advertising clip is composed of four segments, and begins with Kalevipoeg complaining about his new musical instrument falling apart just after the end of the warranty period:

> Kalevipoeg*: Look at this kannel, it's almost new and the one-year warranty just ended yesterday.*

This is followed by the hedgehog's advice in the style of improvisation, in the style of folk poetry, and a neutral style sentence offering information about consumer protection.

> Hedgehog*: Kalevipoeg, don't cast your music into the corner,*
> *take your tune to the toilet!*
> *Your right to submit a complaint to a merchant lasts for two years. Stand up for your rights, see tarbijakaitseamet.ee.*

The fourth phrase is self-referential once again, and returns to Kalevipoeg's failed audition, while the verbal text generalises the message of the advertisement:

> Metsatöll*: One can see that the boy wants; come back as soon as your instrument is in good working order!*

Kalev's chocolate advertisement, *Tastes to Remember...* (2012)[5] also relies on allusions. The uploader of the series of Kalev chocolates (the Epic series) made the following comments on February14, 2012:

> *In an epic bow to our proud stock, Kalev's three traditional chocolates, Kalevipoeg, Linda, and Kalev have been brought together in the Epic series. The new packaging concept was developed by the advertising agency Identity.*[6]

In the clip, the viewer, who is in the position of the primary narrator, is leafing through an old book. In the illustrations, scenes from Kalevipoeg's agrarian life slip by: the eagle flying over the land; Kalevipoeg ploughing; the giant pulling a boat up on the shore; the giant on a ship, pulling shipwrecked people up on the deck; Kalevipoeg lying on the grass chewing on a straw. These pictures emphasise characteristic epic motifs.

In the background, one can hear the epic voiceover, built around folk proverbs and parodies of old proverbs recited in verse; the soundtrack is in the background, played on a traditional instrument, the *kannel*.

> *A man works and then he manages, when he manages, then he rows, softly, that must have been a mistake; if you see a mistake to deplore, come and help, when the man helps, he speaks; when he speaks, he thinks, when he thinks, he thinks for a long time, if it's for long time, then it's with wood; with wood, striking, flat, of course. What can you do, that's how it is, and what it was, is in his mind, and what's in his mind is on his tongue, and it is good. A taste to remember.*

One of the proverbs used in the text is *Tasa sõuad, kaugele jõuad* (If you row softly, you will get far) – a popular new proverb or aphorism; also, some traditional proverbs are used, such as *Kus viga näed laita, seal tule ja aita!* (If you see a mistake to chide, come and help!) (EV 14006) and *Enne mõtle, siis ütle!* (Think before you speak!) (EV 7041).

When he thinks, then he thinks for a long time could be an allusion to the slowness of the stereotypical giant (or of Estonians in general). An analogue can be seen in the anecdote of the 'three slow ones' (three trolls, three Finns, or three Estonians). They ponder something silently for what seems to be an eternity, before making the briefest of comments. Thereupon the third one, disgusted, leaves the company of these 'chatterboxes'.

In the next part of the text, a transfer of content occurs through allusions to the epic, where the hedgehog suggests to the hero that he hit the enemies with the boards of lumber – edgewise, rather than flat.

The text continues with a pair of new proverbs: *Mis meelel, see keelel; mis keelel, on hää* ('What's in the mind, is on the tongue', and, 'What is on the tongue is good'), followed by a generalising advertising slogan from a completely different register. The level of knowledge allows the viewer to enjoy the analogies to well-known sayings by using similar scripts, but these also serve to alter the boundaries of understanding.

FROM CORE TEXTS AND MEDIA BACK TO FOLKLORE AND SOCIAL SPACE

The texts of the preceding advertisement clips are part of the living process of folklore, which includes creating parodies of proverbs, making links between the parodic forms and the original proverbs, generating imitations of alliterative folk poetry (*regivärss*) in an ethnically marked fashion, and using devices derived from folk narrative that inflate or exchange the attributes of characters (for example, replacing the gigantic Kalevipoeg with a delicate young man and representing the folk rock singers as tall as giants). The richness of devices and forms is used to create a cohesive whole point strongly to the existence of

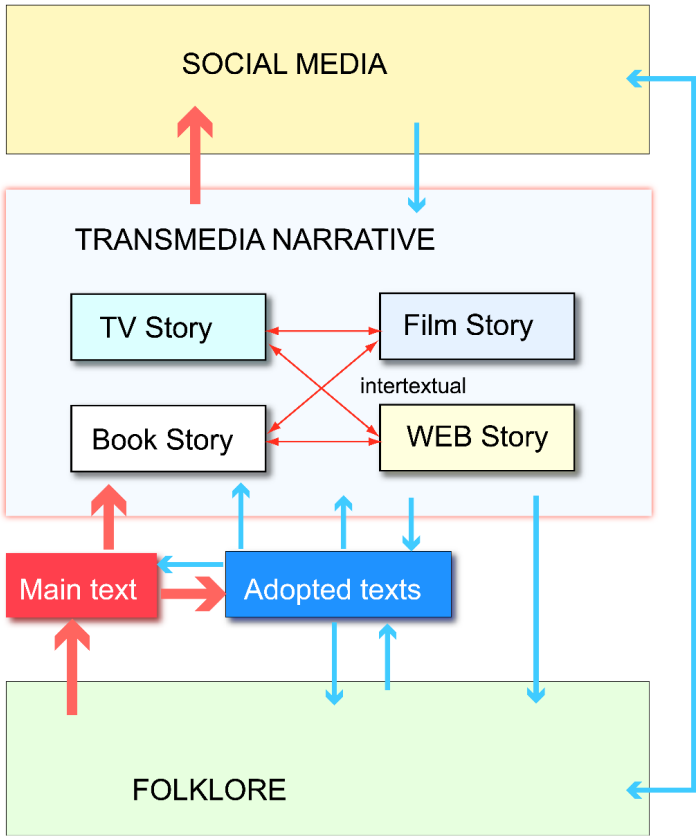

Figure 2. *Folklore and transmedia narrative.*

knowledge, characters, poetics, textual devices, and scripts that are shared in the cultural space; one can draw upon any of these, or parody any one of them. This is also an ideal example of the way that today's media texts amalgamate things, and the wealth of devices that they have at their disposal.

Here it is appropriate to mention another general tendency: the prominence and media success that are achieved when a work or a character causes a flow of various sayings, jokes, and aphorisms back into folklore (on Estonian sayings see, e.g., Peebo 1997).

An excerpt from the 16th tale of the epic *Kalevipoeg* was recirculated into 20th century imagery, the culture of verse albums, photos, and memory albums, thereby acquiring an aphoristic quality:

> *Ülemaks kui hõbevara,*
> *kallimaks kui kullakoormad,*
> *tuleb tarkus tunnistada.*[7]
> Wisdom should be deemed
> Greater than silver
> Dearer than loads of gold.

In figurative speech, *kalev* / *Kalevipoeg* (originally a signifier of physical size and strength, cf. Annist 2005: 307ff.) has come to be a marker of the Estonian, either as an ethnonym or as a basis of comparison; also Tõll was used in figurative speech (Justkui 1998–2005). The same semantic field was evoked in a recent television series with comic overtones, *The* Kalevipoeg*s in Finland*, depicting the lives of construction workers. The expression *Kalevite kange rahvas* (Kalevs' strong nation) is a stock phrase that has spread as a quotation from scholarly texts (cf., e.g., Lang 2012) into everyday speech. Whereas the poet Johannes Semper used this phrase in the national anthem of the Estonian SSR, a present-day NGO that specialises in organising fantasy games has chosen *Kalevite kange rahvas* as its official name.

Expressions from the epic, such as "When all the pine splints burst into flame at both ends", and "Kalev will come home one day, bringing happiness to all his children", have been culturally quoted slogans and sources of allusions both for literary writers and ordinary users. These quotations have also been central in identity-creating documents (the 1918 document, *Manifesto to All the Peoples of Estonia*).

These are only a few examples of the wider circle of dissemination for such expressions in social space, and they can direct our attention to the particular value of mythology and epic at historical junctures critical to establish definition as a nation. Indeed, they were foregrounded during the national movement at the end of the 19th century (cf. Viires 1990), and have continued to be emphasised during the 20th century, holding an important place both during the creation of the Estonian republic and in the years of Soviet annexation.

During both of the abovementioned periods (the interwar republic and the postwar Soviet period), the marking of streets and institutions with signifiers from the national epic was one possible means of national self-assertion. (After the Second World War, the name of the Estonian confectionary company was changed to Kalev, and brands of sweets were designated with names derived from the epic; similarly, the name Tõll was given to an icebreaker ship).

Kalev, Sulev, Olev, and Linda became prominent choices for children's first names; emigrés named their new buildings and associations with ethnonyms and the signifiers of national figures (e.g. Tuisk 2001). Kalevipoeg was used in the rhetoric of speeches on festive occasions and in nationally-oriented consumer texts, both in the homeland and in communities abroad.[8]

One crucial marker of transmedia narratives is the extremely swift crossing of cultural borders by fantasy characters and their transfer from the media into material consumer culture. This is equivalent to the creation of a franchise. According to Jenkins, it is a well-known characteristic of cultural texts that they extend themselves into other cultural forms and into society. For Estonians and experts on Estonian culture, the epic and its protagonist, Kalevipoeg, are matters one must know about and cannot ignore. From the preceding examples it is clear that elements of an epic character or epic text have taken root in different segments of social space. During waves of migration and emigration from Estonia before and after the interwar Estonian republic, the national epic was taken along to faraway places and transplanted in the form of toponyms and group names.

We can assume that in some cases both market uses and the success of franchises are inscribed into certain epics. Examples of this include Kalevipoeg souvenir dolls and various other marketable products. In contemporary culture Kalevipoeg participates very broadly in the entertainment world, from computer games to souvenirs. There are hiking trails, travel programmes, museums, and theme parks connected with Kalevipoeg and Tõll, and the pub-hotel Tõll is a good example of transmedial extensions into the community.

However, in core (and mass media) texts and in their social uses there are implicit aspects pertaining to mentality. Having received a generative impulse from the same basic texts, they reveal their deep connection to group identity and complex social processes.

A setting or filming site may become a cult location, a destination for pilgrimages as designated by new tourism and religiosity. A work may be the starting point for religious groups or movements (cf. *Star Wars*, see also Dawson 2003, Lucas & Robbins 2004), indicating that its problematics are much broader than issues of sales success, the shaping of childhood landscapes, nostalgia, or cultural knowledge shared by multiple generations.

SUMMARY

Taken as an umbrella term under which one can situate folkloric characters, spontaneous dissemination of media texts, and relations between cultural space and social processes, transmediality is indeed an appropriate term for complicated dialogical processes.

> *Transmedia* [---] *implies a source-oriented process whereby a prototext is diverged into different individual metatexts in cultural space. Examples of transmediality would also include cinematic adaptations of a preexisting novel or the spontaneous ways certain motifs move from literature to music to painting, etc., in culture over time*. (Ojamaa 2015)

As an intergenerational core text, the phenomenon of the epic has furnished a wealth of dense material to fill cultural space; in the 20th century elements of the epic have spread from one medium to another. The success of different subtypes and forms has been variable, and they have had recourse to a range of stylistic and formal resources. In the wake of the first printings of these texts, adaptations quickly emerged, facilitating the broadly-based, educational appropriation of the epic material. The creators of retellings and adaptations[9] have included well-known literary figures such as Juhan Kunder (1885), Villem Ridala (1921), Eduard Laugaste (1960), Eno Raud (1961), and others.

In fact, Eno Raud's prose narratives about Kalevipoeg and Tõll were so popular that they went through many reprints, and they were translated into Latvian, Ukrainian, Russian, Finnish, and German. According to the data of the National Library, in the years 1857–2002 there have been 17 reprints of the epic.

Nevertheless, adaptations weaken the canon, and it gradually becomes easier to transgress its boundaries emotionally and intellectually, resulting in parodies. In Estonian media space, parodies began to circulate later than they had in the realm of folklore. It was not until the 1970s (e.g. Vetemaa 1971; Rakke 2000; Kirsfeldt 2010; cf. Laitila 2003) that the professional media began to substitute antiheroes or ordinary people for heroes, thus arriving at the form of caricature.

The examples of Kalevipoeg and Tõll indicate that in order to enter the media, a folkloric corpus requires an author who creates and gives shape to a specific tale – or, using the modern term, a storyworld.

There are far more thematically promising (perhaps even better?) motifs and elaborations in folklore than an author could possibly actualise, and these await their time in a state of latency. Media transfers from a work and the success thereof depend on a certain intracultural expectation. Indeed, there was already a certain preliminary social and cultural subscription to the figures of Kalevipoeg and Tõll.

The dissemination of Kalevipoeg and Tõll in transmedia and culture has been multifaceted. In today's media space it is complicated to differentiate the epic *Kalevipoeg* from the folkloric Kalevipoeg, as well as variants of the character that have been spread through other texts in the cultural space. As such, the character is an amalgam of different sources constituting new generalisations. The same conclusion applies to Tõll, whose range of social impact has been more local from the beginning and its spread more limited than that of Kalevipoeg; in today's terms, the character of Tõll is a fusion of professional art and folklore.

Unelaborated (raw) folkloric texts about giants appeared in print relatively late and after longer intervals; they have entered the cultural circulation, often as mediators of other material. However, there is certainly a firm and enduring place for new media texts and consumer texts such as advertising clips, as the folkloric process embraces their multiplicity.

ACKNOWLEDGEMENTS

Research for this article was supported by the Centre of Excellence in Estonian Studies (CEES, European Regional Development Fund) and is related to research project IUT 22-5 (Estonian Research Council). The author wishes to thank the anonymous reviewers for the valuable comments.

NOTES

[1] The most recent of these translations have been published in the last five years: the translation into English in 2011, into Hindi in 2012.

[2] The biggest and oldest confectionary company in Estonia.

[3] See https://www.youtube.com/watch?v=--i3pwGGabE.

[4] See https://www.youtube.com/watch?v=iOenRu7mqTQ.

[5] It is worth mentioning that today there is a similar commercial for Suur Tõll chocolate (2016, https://www.youtube.com/watch?v=c9xw52kGKNU), "choclate full of large peanuts".

[6] See https://www.youtube.com/watch?v=roFKVzvGdy4.

[7] Arvo Krikmann (1999) regards this expression as a biblical influence, though he acknowledges that it "has acquired oral circulation, later supported by school textbooks and other printed sources". The use of folklore in verse albums and aphorisms written on the backs of photographs was widespread in the 20th century.

[8] This is a topic for a longer, independent investigation.

[9] The National Library's list indicates that there are 21 works as of 1998.

REFERENCES

Annist, August 2005. *Friedrich Reinhold Kreutzwaldi "Kalevipoeg"*. [Kalevipoeg by Friedrich Reinhold Kreutzwald.] Tallinn: Eesti Keele Sihtasutus.

Baran, Anneli 2008. Fraseologismide semantilise teisititõlgendamise võimalikkusest. [On the Semantic Ambiguity of Phraseologisms.] In: Helle Metslang & Margit Langemets & Maria-Maren Sepper (eds.) *Eesti Rakenduslingvistika Ühingu aastaraamat // Estonian Papers in Applied Linguistics* 4. Tallinn: Eesti Rakenduslingvistika Ühing, pp. 17–33. Available at http://www.rakenduslingvistika.ee/ajakirjad/index.php/aastaraamat/issue/view/ERYa4, last accessed on June 4, 2016.

Dawson, Lorne L. (ed.) 2003. *Cults and New Religious Movements: A Reader*. Blackwell Readings in Religion. Malden & Oxford & Melbourne & Berlin: Blackwell Publishing.

Eisen, Matthias Johann 1901. *Endised jõumehed: Lood Kalewipojast, Suurest-Töllust, Leigrist ja teistest*. [Former Strongmen: Stories of Kalevipoeg, Suur-Töll, Leiger, and Others.] Narva: R. Põder.

Eisen, M[atthias] J[ohann] 1910. *Kalewipoja esi-isad*. [Forefathers of Kalevipoeg.] Tartu: Hermann.

Eisen, Matthias Johann 1913. *Eesti muistsed jumalad ja vägimehed*. [Estonians' Ancient Gods and Mighty Men.] Tartu: G. Roht.

Eisen, Matthias Johann 1920. *Teised Eesti kohalikud muistejutud*. [Other Estonian Local Legends.] Tallinn: Eestimaa Kooliõpetajate Wastastikku Abiandmise Selts.

Eisen, Matthias Johann 1924. Kalevipoja sängid. [Beds of Kalevipoeg.] *Eesti Kirjandus*, No. 4, pp. 143–153, available at http://www.digar.ee/arhiiv/nlib-digar:53209; No. 5, pp. 202–210, available at http://www.digar.ee/arhiiv/nlib-digar:53210, both last accessed on June 4, 2016.

Eisen, Matthias Johann 1926a. Peata Tõll. [Headless Tõll.] *Eesti Kirjandus*, No. 1, pp. 40–44. Available at http://www.digar.ee/arhiiv/et/perioodika/26684, last accessed on June 4, 2016.

Eisen, Matthias Johann 1926c. *Kalewipoja album. A. Adamsoni, O. Hoffmanni, A. Prometi ja A. Siwardi pildid*. [Kalevipoeg Album: Book of Reproductions.] Tartu: Eesti Kunstikirjastus Konze.

Eisen, Matthias Johann 1927. *Tõll ja ta sugu*. [Tõll and His Kin.] Tartu: Eesti Kirjanduse Selts. Available at http://www.digar.ee/arhiiv/et/raamatud/15226, last accessed on June 4, 2016.

Eisen, M[atthias] J[ohann] 1930. Kalevipoja kinnismuistised. [Ancient Relicts Connected with Kalevipoeg.] In: Fr[iedrich] R[einhold] Kreutzwald. *Kalevipoeg. Eesti kangelaseepos*. Parandatud ja illustreeritud väljaanne koolidele. Tallinn: Töökool, pp. 217–221.

EV = Krikmann, Arvo & Sarv, Ingrid (eds.) 1980–1988. Eesti vanasõnad, Vols. 1–4. [Estonian Proverbs.] Tallinn: Eesti Raamat.

Gaston, Kalle 1999. *Tarvastu jutud 2.* [Stories from Tarvastu Parish.] Põltsamaa: Vali Press.

Hupel, August Wilhelm 1782. *Topographische Nachrichten von Lief- und Ehstland*, Vol. III. Riga: Johann Friedrich Hartknoch.

Jenkins, Henry 2004. The Cultural Logic of Media Convergence. *International Journal of Cultural Studies*, Vol. 7, No. 1, pp. 33–43. http://dx.doi.org/10.1177/1367877904040603.

Jenkins, Henri 2006. *Convergence Culture: Where Old and New Media Collide.* New York: New York University Press.

Jenkins, Henri 2007. *Transmedia Storytelling 101.* Available at http://www.henryjenkins.org/2007/03/transmedia_storytelling_101.html, last accessed on June 3, 2016.

Joonuks, Helmut 1982. *Kalevipoja paigad.* [Places Connected with Kalevipoeg.] Tallinn: Eesti Raamat.

Justkui 1998–2005 = *Justkui ehk eesti kõnekäändude ja fraseologismide andmebaas.* [Justkui: Database of Sayings and Phraseology.] Compiled by Anneli Baran, Anne Hussar, Asta Õim, Katre Õim. Available at http://www.folklore.ee/justkui/, last accessed on June 3, 2016.

Kirsfeldt, Kristian 2010. *Kalevipoeg 2.0: Eesti rahva moodne eepos.* [Kalevipoeg 2.0: Modern Epic of Estonians.] Tallinn: Eesti Ajalehed.

Kivirähk, Andrus 2014. *Suur Tõll.* [Tõll the Giant.] Tallinn: Varrak.

Krikmann, Arvo 1999. *Eesti lühivormide allikaloost.* [Research History of Estonian Short Forms.] Tartu: Eesti Kirjandusmuuseum. Available at http://www.folklore.ee/~kriku/ALLIK/laiindex.htm, last accessed on June 3, 2016.

Kunder, Juhan 1885. *Kalevipoeg: Lugu Eesti muinaskangelasest.* [Kalevipoeg: Story about Estonian Ancient Hero.] Tartu: H. Laakmann.

Käo, Henno 2001. *Maailma loomine ja teisi lugusid.* [World Creation Myths and Other Stories.] Tallinn: Ilo.

Körber, Martin 1899. *Oesel einst und jetzt.* Zweiter Band. *Die Kirchspiele Mustel, Kielkond, Anseküll, Jamma, Wolde und Pyha.* Arensburg: Arensburger Wochenblattes.

Laitila, Anu 2003. Kangelase metamorfoos: rahvuseeposte ["Kalevala" ja "Kalevipoeg"] rahvuslikud paroodiad. [The Hero's Metamorphosis: Ethnic Parodies of National Epics *Kalevala* and *Kalevipoeg*.] In: *Paar sammukest XX: Eesti Kirjandusmuuseumi aastaraamat*. Tartu: Eesti Kirjandusmuuseum, pp. 121–128.

Lang, Valter 2012. Kalevite kange rahvas. [Kalev's Strong Nation.] *Keel ja Kirjandus*, No. 12, pp. 877–888. Available at http://www.digar.ee/viewer/et/nlib-digar:114730/141991/page/7, last accessed on June 3, 2016.

Laugaste, Eduard 1960. *Kalevipoeg*. Ümber jutustanud E. Laugaste. [Kalevipoeg: Retold by E. Laugaste.] Tallinn: Eesti Riiklik Kirjastus.

Laugaste, Eduard & Liiv, Ellen 1970. *Muistendid Vanapaganast.* [Legends about Old Nick.] Eesti muistendid: Hiiu- ja vägilasmuistendid III. Monumenta Estoniae Antiquae II. Tallinn: Eesti Raamat.

Laugaste, Eduard & Liiv, Ellen & Normann, Erna 1963. *Muistendid Suurest Tõllust ja teistest* [Legends about Suur Tõll and Others]. Eesti muistendid: Hiiu- ja vägilasmuistendid II. Monumenta Estoniae Antiquae II. Tallinn: Eesti Riiklik Kirjastus.

Laugaste, Eduard & Normann, Erna 1959. *Muistendid Kalevipojast.* [Legends about Kalevipoeg.] Eesti muistendid: Hiiu- ja vägilasmuistendid I. Monumenta Estoniae Antiquae II. Tallinn: Eesti Riiklik Kirjastus.

Leppik, Eduard 2001. *Kalevipoja hobuse jäljerida.* [Traces of Kalevipoeg's Horse.] Rakvere: Shalom.

Lincoln, Bruce 2000. *Theorizing Myth: Narrative, Ideology, and Scholarship.* Chicago & London: The University of Chicago Press.

Lucas, Philip Charles & Robbins Thomas 2004. *New Religious Movements in the Twenty-First Century: Legal, Political, and Social Challenges in Global Perspective*. New York & London: Routledge.

Luce, Johann Wilhelm Ludwig von 1827. *Wahrheit und Muthmassung: Beytrag zur ältesten Geschichte der Insel Oesel*. Pernau: Gotthardt Marquardt.

Moloney, Kevin T. 2011. *Porting Transmedia Storytelling to Journalism.* A Thesis Presented to the Faculty of Social Sciences, University of Denver. Available at http://www.colorado.edu/journalism/photojournalism/Transmedia_Journalism.pdf, last accessed on June 3, 2016.

Moloney, Kevin 2011–2014. *Transmedia Journalism.* Available at https://transmediajournalism.org/contexts/, last accessed on June 3, 2016.

Niineste, Mart 2010. Suur Tõll 30 aastat hiljem. [Suur Tõll 30 Years Later.] Newspaper *Eesti Päevaleht*, 23.01. Available at http://epl.delfi.ee/news/kultuur/suur-toll-30-aastat-hiljem?id=51186549, last accessed on June 3, 2016.

Ojamaa, Maarja 2015. *The Transmedial Aspect of Cultural Autocommunication.* Dissertationes Semioticae Universitatis Tartuensis 19. Tartu: University of Tartu Press. Available at http://dspace.ut.ee/bitstream/handle/10062/45985/ojamaa_maarja.pdf?sequence=1, last accessed on June 3, 2016.

Peebo, Kadri 1997. "If I'm Not Mistaken, But I'm Not Mistaken, If I'm Not Mistaken by Any Chance": Catchwords? Quips? Bywords? *Journal of the Baltic Institute of Folklore*, Vol. 2. Available at http://www.folklore.ee/rl/pubte/ee/bif/bif2/sisu.html, last accessed on June 4, 2016.

Raamat, Rein 1982. *Suur Tõll*. [Tõll the Giant.] Tallinn: Perioodika.

Rakke, Kerttu 2000. *Kalevipoeg*. Tallinn: Kadikas.

Raud, Eno 1961. *Kalevipoeg*. Tallinn: Eesti Riiklik Kirjastus.

Ribenis, Karin (ed.) 2002. *Eesti rahvaluule elektrooniline bibliograafia (1918–1992) / Electronic Bibliography of Estonian Folklore (1918–1992)*. Tartu & Tallinn: Eesti Kirjandusmuuseumi folkloristika osakond. Available at http://www.folklore.ee/rl/pubte/ee/erbibl/, last accessed on June 3, 2016.

Ridala, Villem 1921. *Kalevipoeg*: *[proosaümberjutustus].* [Kalevipoeg: Prose Narration.] Tallinn: Kool.

Ringvaade 22.10.2014. *Kivirähk ja Arrak avaldasid uue "Suure Tõllu".* [Kivirähk and Arrak Published a New "Suur Tõll".] Available at http://etv.err.ee/v/meelelahutus/ringvaade/ringvaate_lood/e9dd6b3b-f992-48fc-bbcf-7299d8528f8d, last accessed on June 4, 2016.

Ryan, Marie-Laure 2008. Transfictionality across Media. In: John Pier & José Angel Garcia Landa (eds.) *Theorizing Narrativity*. Berlin: De Gruyter, pp. 385–417.

Sonnenfeld, Gunther 2009. *The Revolutionary Power of Transmedia Storytelling.* September 10. Available at https://www.imediaconnection.com/articles/ported-articles/red-dot-articles/2009/sep/the-revolutionary-power-of-transmedia-storytelling/, last accessed on June 3, 2016.

Süda, Peeter 1883. *Väikene vana varanduse vakk ehk Saaremaa vägimees Suur-Tõll. Eestlaste ennemuistne jutt. Saaremaa rahva suust 20-ne aasta jooksul kokku korjatud.* [Small Old Treasure Chest or Saaremaa Strongman Suur Tõll.] Kuressaare.

Tedre, Ülo 2003. *Eesti mees ja tema sugu XIX sajandi lõppriimilises rahvalaulus.* [Estonian Man and His Kin.] Tallinn & Tartu: Eesti Kirjandusmuuseum.

Tuisk, Astrid 2001. Siberi eestlaste kohapärimus tänapäeval: Kohastumise peegeldusi. [Place Names of Contemporary Siberian Estonians: Reflections of Adaptation.] In: Tiiu Jaago (ed.) *Pärimuslik ajalugu.* Tartu: Tartu Ülikooli Kirjastus, pp. 74–90.

Vetemaa, Enn 1971. Kalevipoja mälestused. [Memoirs of Kalevipoeg.] *Looming*, No. 11, pp. 1619–1648, 1651–1671, available at http://www.digar.ee/arhiiv/et/perioodika/8480; No. 12, pp. 1763–1808, 1811–1820, available at http://www.digar.ee/arhiiv/et/perioodika/8481,both last accessed on June 3, 2016.

Viires, Ants 1990. Taara avita! [Taara, Help!] *Looming*, No. 10, pp. 1410–1421. Available at http://www.digar.ee/arhiiv/et/perioodika/8719, last accessed on June 4, 2016.

YOUTUBE CLIPS

Metsatöll: *Vaid vaprust.* [Metsatöll's music clip *Only Bravery*.] 2010. Available at https://www.youtube.com/watch?v=--i3pwGGabE, last accessed on June 4, 2016.

Kalevipoeg teeb ansamblile Metsatöll prooviesinemist. [Kalevipoeg Performing for the Metsatöll Band.] 2012. Available at https://www.youtube.com/watch?v=iOenRu7mqTQ, last accessed on June 4, 2016.

Maitsed mis on meeles. [Tastes to Remember.] 2012. Available at https://www.youtube.com/watch?v=roFKVzvGdy4, last accessed on June 4, 2016.

Suur Tõll tume šokolaad. [Suur Tõll Dark Chocolate.] 2016. Available at https://www.youtube.com/watch?v=c9xw52kGKNU, last accessed on June 4, 2016.

OUR AUTHORS

Renate Pajusalu is Professor of General Linguistics at the Institute of Estonian and General Linguistics, University of Tartu, Estonia. Her research interests include semantics and pragmatics of Estonian and related languages from cognitive and interactional perspective, and first and second language acquisition.

renate.pajusalu@ut.ee

Anni Jürine is specialist-assistent at the Centre for Academic Writing and Communication, University of Tartu, Estonia. Her main fields of research include language change, especially grammaticalization and lexicalization, as well as academic writing.
anni.jyrine@ut.ee

Karl Pajusalu is Professor of History and Dialects of the Estonian Language at the Institute of Estonian and General Linguistics, University of Tartu, Estonia. His main areas of research are phonological and grammatical variation of Estonian and related languages, and historical formation of the Estonian language area.

karl.pajusalu@ut.ee

Ilona Tragel is a linguist. She is an associate professor of general linguistics at the University of Tartu, Estonia, and a lecturer of Estonian language and culture at the Beijing Foreign Studies University in Beijing, China. Her main fields of research are semantics and grammaticalisation of frequent verbs in the framework of cognitive linguistics.

ilona.tragel@ut.ee

Ann Veismann is Research Fellow in general linguistics at the University of Tartu, Estonia. Her research relates to cognitive linguistics, and her main research interests are semantics of adpositions and adverbs, space and time expressions in Estonian, and empirical methods in semantics.

ann.veismann@ut.ee

Yuri Berezkin is Head of the American Department of the Peter the Great Museum of Anthropology and Ethnography (Kunstkamera), Russian Academy of Sciences, and Professor at the Department of Anthropology, the European University at Saint Petersburg.

berezkin1@gmail.com

Evgeny Duvakin is an independent researcher and a participant in the "Stories Retold over Millennia: Reconstruction of the Dynamics of the Global Distribution of Replicated Elements of Oral Narratives" research project of the Peter the Great Museum of Anthropology and Ethnography (Kunstkamera), Russian Academy of Sciences.

e.duvakin@gmail.com

Mari Uusküla holds a Ph.D. in linguistics and is an Associate Professor of Translation Studies at Tallinn University. Her research interests include linguistics, lexical semantics, semantic typology, color naming, categorization and perception, field linguistics, and psycholinguistics.

mari.uuskula@tlu.ee

David Bimler holds a BSc in physics and a Ph.D. in mathematical psychology. He is a Research Associate with Massey University in New Zealand. His research focuses on vision, considering such topics as the perception of color and facial expressions of emotion.

d.bimler@massey.ac.nz

Jodi Louise Sandford is a researcher and adjunct professor in English language and linguistics at the University of Perugia, Italy. Her research in cognitive semantics has concentrated on seeing or color words, categorization, and conceptualization.

jodi.sandford@unipg.it

Natalia Glukhova Chair of Finno-Ugric and Comparative Philology, Institute of National Culture and Intercultural Communication, Mari State University.

gluhnatalia@mail.ru

Hille Pajupuu, researcher at the Institute of the Estonian Language, Tallinn, Estonia. Fields: speech perception and speech acoustics.

eki@eki.ee

Rene Altrov, researcher at the Institute of the Estonian Language, Tallinn, Estonia. Fields: linguistic corpora, intercultural communication.

eki@eki.ee

Jaan Pajupuu, software developer.

eki@eki.ee

Liisi Piits is researcher at the Institute of the Estonian Language. Her main fields of research include collocational behaviour of nouns and phonological variation of Estonian pronunciation.

Liisi.Piits@eki.ee

Piret Voolaid is senior researcher at the Department of Folkloristics, Estonian Literary Museum. Her current interests include subgenres of Estonian riddles, proverbs in their various contemporary contexts, proverbs, children's and youth, Internet, and sports lore.

piret@folklore.ee

Ekaterina Velmezova is linguist, Slavist, full professor of Slavistics and of history and epistemology of language sciences in Eastern Europe at the University of Lausanne, Switzerland. Her main fields of research include history and epistemology of language sciences, and Slavic ethnolinguistics.

ekaterina.velmezova@unil.ch

Mare Kõiva is folklorist and leading researcher at the Department of Folkloristics, Estonian Literary Museum, Tartu, Estonia. Her main fields of research include mythology, belief narratives and incantations, religious studies and new spiritual movements, ritual year, contemporary folklore, relations between humans and animals.

mare@folklore.ee

Andres Kuperjanov is the folklorist and editor at the Estonian Literary Museum, Tartu, Estonia. His main fields of research include archeoastronomy and ethnoastronomy, popular calendar, ritual practices, belief narratives, new religious movements, digitalhumanities.

cps@folklore.ee

Dedicated to Urmas Sutrop

vol. 64 *http://www.folklore.ee/folklore*

On the cover: Eduard Philipp Körber's drawing from 1800 of a wooden figure of God attributed to pagan Estonians, with a staff symbolising Christian power (J. Ch. Brotze's collection, Riga, Latvia).

9 771406 095006

Folklore EJF

ISSN 1406-0957