

The article is published in *Diachronica*, 40(4), 492-531 (2023).

DOI: <https://doi.org/10.1075/dia.22013.boc>

**Linguistic mechanisms of colour term evolution:
A diachronic investigation of 'Russian browns' *buryj* and *koričnevij***

Vladimir V. Bochkarev¹, Anna V. Shevlyakova¹, Valery D. Solovyev¹, Ekaterina V. Rakhilina²
and Galina V. Paramei³

¹ Kazan Federal University, ² Higher School of Economics University, ³ Liverpool Hope University

We investigated diachrony of distributional semantics of two competing Russian colour terms (CTs) for 'brown', *buryj* (11th century) and *koričnevij* (17th century), using the Russian subcorpus of Google Books Ngram (2020). Time-series analysis (1800–2019) of bigrams gauged each term's frequencies of occurrence and changes in combinability with nouns for natural objects, artefacts, abstract concepts and figurative expressions. In frequency, *koričnevij* overtook *buryj* in the 1920s, confirming its basic status in modern Russian. The perplexity index indicates that *koričnevij* steadily increased the range of denoted objects, with artefacts being front runners in the *buryj*-to-*koričnevij* transition. The results corroborate Rakhilina's (2007a, 2007b, 2008) hypothesis that an incipient CT initially collocates with nouns denoting artefacts but gradually expands to the realm of natural objects supplanting an old CT. Moreover, *koričnevij* and *buryj* are discerned by denotations and connotations. The present findings provide insights into general mechanisms of the linguistic evolution of an emergent basic CT.

Keywords: "Russian browns", computational linguistics, Google Books Ngram, colour term collocational dynamics, diachronic distributional analysis, linguistic colour term evolution

1. Introduction

In the present study we address the problem of the linguistic evolution of basic colour terms (BCTs) by focussing on two Russian terms for ‘brown’, an old term *buryj* ‘dust/greyish brown’, and a new term *koričnyj*, which is an established BCT in modern Russian. We undertook a diachronic computational analysis of Google Books Ngram to investigate the contextualised linguistic behaviour of each term to explore their combinatorial lexical typology and, specifically, the semantic (collocational) dynamics of the new term. In our endeavour to understand the puzzling case of *buryj* being supplanted by *koričnyj*, we aspire to provide general insights into the driving forces and mechanisms – perceptual, cognitive, linguistic, and socio-cultural – of BCT evolution.

We start with an overview of the explanatory schemes of BCT evolution as proposed by Berlin, Kay and colleagues (Berlin & Kay 1969/1991; Kay & Maffi 1999; Kay 2015), as well as MacLaury’s (1991, 1997) elaborations. We then proceed with historical-linguistic information on the origin, emergence and meanings of the two “Russian browns”. Unlike more traditional analysis of a term’s denotative meaning, informed by using a decontextualised methodology, we expound the term’s meanings and connotations through an analysis of their contextualised linguistic behaviour. The Introduction is concluded with the present study’s hypothesis.

1.1 The Berlin-Kay hypothesis of basic colour categories

Colour, as percept, or qualium, is a product of human psychology as well as physiology and genetics. As a concept, in comparison, colour is closely linked to language: it reflects the requirements of cognitive economy and effective communication, to partition the gamut of all perceived colours into a certain number of colour categories, and to label each category with a term (Conway et al. 2020; Zaslavsky et al. 2020, 2022; Chaabouni et al. 2021).

The Sapir-Whorf linguistic relativity hypothesis posits that divisions of the colour continuum are arbitrary and differ between languages (Saunders & van Brakel 1997). In opposition to this view is the Berlin and Kay (1969/1991) hypothesis of pan-human basic colour categories (BCCs) and the corresponding BCTs, currently dominant in colour cognition research. In the context of the present study, it is worth noting differences in methodological approaches of the two views’ proponents: advocates of the relativity hypothesis pursue analysis of colour words using linguistic methods; the universalist view, in comparison, was

developed from a psycholinguistic perspective, such that the meaning of a colour word is validated denotatively, by specific referents from an array of colour samples.

Relevant to the present study are two aspects of the Berlin-Kay hypothesis, described here as given in 1969/1991:

- a. To be considered basic, a colour term should meet certain criteria (Berlin & Kay 1969/1991: 5–7). Among these, some are linguistic: a BCT is monolexemic; it is a non-hyponym; and it collocates with nouns that are not restricted to certain classes of objects. The other criteria are psychological: a BCT is cognitively salient, and understood by all speakers of the language.
- b. BCCs/BCTs recur across languages and evolve in a partially fixed order, from a minimum of 2 (stage I) to a maximum of 11 (stage VII; in industrialised societies). Individual languages are classified as being at a certain stage in the colour terminology expansion process, “having” a certain number of BCTs. By stage V, a language possesses all six primary BCTs (‘white’, ‘black’, ‘red’, ‘green’, ‘yellow’ and ‘blue’). The term for ‘brown’, one of the secondary BCTs, emerges at stage VI. At the final stage VII, ‘orange’, ‘pink’, ‘purple’ and/or ‘grey’ are added.

The Berlin-Kay hypothesis makes diachronic predictions about the emergence of a new BCC/BCT, in that languages move from one stage in the sequence to another. The hypothesis triggered enquiries about the trajectories of incipient BCCs and, importantly, about the mechanisms of development of new colour constructs and the driving forces of the construct lexicalisation. A recent analysis provided evidence that the BCT evolution pattern may vary among languages: along with the colour term gain, the loss of at least some colour terms has also been observed in the process of the interaction of cognitive constraints and language change in shaping lexical systems (Haynie & Bower 2016).

The enquiries also sparked criticisms on linguistic grounds. Note that, regardless of the angle of the criticisms, they tacitly assume that during the evolution of colour terms, there exists a one-to-one correspondence between an incipient BCC (a concept) and the corresponding BCT (a name). This assumption is challenged in the present study. To put this into context, we review the current explanatory schemes and mechanisms of BCC evolution.

1.2 Evolution of basic colour categories as partitioning of colour space:

Explanatory schemes and mechanisms

Berlin and Kay (1969/1991) assumed that in all languages, the observable colour gamut is partitioned by the available BCCs/BCTs, the idea epitomising Werner's (2004) general theory of development by differentiation. Across languages, two partition scenarios of BCCs are observed: (i) a novel BCC can emerge as a result of an "old", composite, BCC splitting into two (exemplified by a fission of 'yeen' into 'yellow' and 'green'); (ii) a new category can be inserted that straddles the boundaries of existing BCCs (exemplified by 'orange' inserted between 'red' and 'yellow').

The partition of colour space by BCCs, and the corresponding augmenting of the BCT inventory, are conceived to unfold as a way of meaningful information coding of the colour gamut in order to guarantee the communicative success of speakers; the process implies an interaction of perceptual and cognitive factors. Accordingly, BCCs are added so that they maximise colour differences between adjacent categories and minimise colour differences within the new categories (the Optimality Hypothesis; Jameson & D'Andrade 1997). A computational simulation confirmed the optimality criterion of colour space partitioning as the explanatory scheme of the evolution of up to six primary BCCs (Regier et al. 2007). However, the simulation beyond the primary BCCs failed to reproduce the empirically obtained data (Jraissati & Douven 2017). The authors interpreted the failure as evidence that, along with the optimality criterion, colour conceptualisation requires an explanation in broadly cultural terms. Jraissati and Douven (2017) conjecture that the development and shaping of novel (basic) CCs is driven and determined by an interaction between linguistic factors, cultural needs and environmental factors: in particular, innovations in technology. This is supported by a diachronic analysis (of an individual language) by Zaslavsky et al. (2022), who conclude that an increasingly fine-grained colour-naming system is driven by a functional need for communication to be both accurate and simple.

1.3 Evolution from a "proto-archaic" to a new hue-based basic colour category and term

Relevant to the evolutionary dynamics of "Russian browns" is an alternative scheme of emerging BCCs put forward by MacLaury (1991, 1992, 1997). Based on Greenfeld's (1986) observations and his own findings in the Mesoamerican Color Survey, MacLaury concluded that, in denotative terms, desaturated-complex categories that engulf the grey "core" of colour space are evolutionarily early. These categories have basic status and "emerge on the least saturated but blended areas of the solid's surface, commonly brown, lavender, beige

and pale” (MacLaury 2007: 125). Under specific cultural pressures, such “proto-archaic”, brightness-based BCCs evolve through composite brightness-and-hue categories, to the gradual crystallisation of hue-only categories (for further details see Supplement S1). Plausibility of the brightness-to-hue trajectory was recently demonstrated in computer simulations (see Supplement S2).

MacLaury (1992) offers answers to three questions arising from his scheme. As a *motivation* for such conceptual dynamics, he suggests novelty and innovation in a speaker’s community, including new technological inventions. As an underlying *cognitive mechanism*, he conjectures that there would be an increasing emphasis on distinction, rather than similarity, between category members. *Linguistically*, the term for the incipient category, as a rule, is a borrowing from a contact language that possesses a term for a referent which acts as an exemplar of the colour concerned.

1.4 Evolution of basic colour terms: A linguistic approach

The hypotheses of BCC emergence based on analyses of denotative meanings were criticised by linguists, who pointed out that the evolutionary dynamics of colour *terms* – as denotation of semantic universals – cannot be comprehended only through decontextualised denotative meanings of colour, and argue that an analysis on a language-internal basis is more compelling (Lyons 1995; Lucy 1997; Gage 1999; Levinson 2000).

Beyond answers to the questions of which colour space areas are prone to the BCC evolutionary dynamics (“where”), and what the origins of new colour terms may be (“wherefrom”), the linguistic approach may provide further insights into the innate mechanisms and driving forces underlying the ways of re-structuring the colour categorisation system (“how”) (see a review by Paramei & Bimler 2021).

As argued by Decock (2021), the linguistic approach helps to capture not only the optimisation of the BCC evolutionary dynamics, but, importantly, the drift of colour concepts over time, as mediated by socio-cultural processes. This view is echoed by Saunders and van Brakel (1997), who posit that cultural processes are crucial for exerting non-trivial constraints on the emergence and development of CTs. Leaning upon their conjecture, Levinson (2000) reiterates the need to investigate how colour vocabulary in a particular language is linked to local practices, technology and aesthetics. Further socio-cultural incentives that transpire in

colour terminology are pragmatic and semantic distinctions (Steels 2011), and cultural symbolism and values (Gage 1999). High value accorded to a particular colour can be the result of the workmanship involved in the coloured objects and/or the distances from which pigments or dyeing materials have been obtained (Gage 1999). In the discussion (§4), we address socio-cultural, economic, historical and implicated value (signifier) factors that, we argue, stipulated the denotative and connotative distinction of the two “Russian browns” and their linguistic evolution.

Levinson (2000: 7) suggested several lines of linguistic enquiry into CTs to properly define the colour term domain: (a) morphology, derivational and collocational possibilities and the structural semantics of word classes denoting colour; (b) “the typical use and full referential range of each expression ... how and in what contexts color ... contrasts are made”; (c) “adequate sampling of the linguistic form classes and semantic fields”. Levinson (2000) appeals for scrutiny of the patterns of CT co-occurrences, i.e. the general practice of linguistic typology. In a similar vein, Wierzbicka (2005) points to the need to explore the combinability of lexemes (here: colour terms) in different contexts, i.e. their linguistic behaviour that is motivated by the terms’ semantic properties and can provide clues to understanding the terms’ meanings – their intensions, rather than denotations or extensions.

1.5 Corpus analysis of the linguistic behaviour of colour terms

An analysis of the diagnostic contexts of a given word (here: a colour term) is attained by corpus analyses within the lexical-combinatorial approach (Koptjevskaja-Tamm et al. 2015). Significant progress in computational linguistics now enables the use of statistics to search for the collocations of a term to identify and group its contexts. Importantly, the computerised corpus analysis of an individual language enables the tracing of diachronic changes in word meaning driven by external factors (cultural, societal and technological), as well as internal language motivations (Gulordava & Baroni 2011; Kulkarni et al. 2015; Tang 2018; Tahmasebi et al. 2021). Corpus-based studies provide evidence that these changes are manifested by the frequency of word usage (Joula 2003; Hilpert & Gries 2009, 2016). As pointed out by Decock (2021: 15), “[i]f statistical changes in the co-occurrences of [colour] concepts within sentences of a larger corpus can be observed over time, these changes will be highlighted in distributional semantics”.

1.6 “Russian browns”: Etymology, history of emergence and linguistic behaviour

In modern Russian, the BCT for ‘brown’ is *коричневый* / *koričnevyyj*, as attested by numerous linguistic and psycholinguistic studies (Berlin & Kay 1969/1991; Hill 1972; Frumkina 1984; Frumkina & Mikhejev 1996; Kul’pina 2001, 2007; Rakhilina 2007a, 2007b, 2008; Vasilevich et al. 2016). Along with this term, Russians name a significant range of objects exclusively with the old term *бурый* / *buryj* ‘dusty/greyish brown, reddish-brown, brownish black’ (Hill 1972; Falla 1984; Krivko 2021).

Apart from the two principal terms for ‘brown’, Russian possesses several old domain-specific terms denoting objects of various brown shades. In particular, *karij* is predominantly used for naming the colour of brown eyes; although less frequently, it also collocates with words for hair, horses, and beavers. Moreover, multiple old terms denote specific colours of horses’ coats, such as *gnedoj* ‘bay/chestnut’, *bulanyj* ‘light bay/buckskin’, *karakovyj* ‘dark bay/black dun’, *kauryj* ‘light chestnut’, *savrasyj* ‘sorrel’, *čalyj* ‘roan’, *muxortyj* ‘bay with yellowish markings’ and others.

Here we focus solely on the two principal Russian ‘brown’ terms. As Corbett and Morgan (1988: 61) concluded in their extensive linguistic study, “the two terms together fill the brown slot in the [Berlin-Kay] hierarchy”. In a psycholinguistic study, the focal colours of *koričnevyyj* revealed double-focality (instability) of *koričnevyyj* “best exemplars” in colour space (Safuanova & Korzh 2007: Figure 3), which hints that one of the two foci might represent the “best exemplar” of a *buryj* denotative slot.

Studies in historical linguistics show that the two ‘brown’ terms differ in the time of their emergence and lexical origin. The old *buryj* apparently emerged in Russian at stage VI of the Berlin-Kay evolutionary sequence, as in other languages. According to Herne (1954), *buryj* is related to Persian *bōr* ‘red; colour of the pistachio nut’ and Turkish *bur* ‘fox-red’, and is the cognate of Mongolian *bürüj* ‘dark-coloured’.

Bakhilina (1975: 40) conjectures that in Old Russian, *buryj* may have been used in everyday language and folklore as one of the colour names for the coats of horses and cattle as early as the 11th or 12th centuries, although written (trade) sources from that period do not attest the use of any colour terms (cf. also Herne 1954). The earliest examples of expressions containing *buryj* are recorded in birchbark letters from the 14th century (Gippius 2020); also, they are attested in Old Russian dictionaries of the 11th–17th centuries (Slovari 11–17 vekov 1975–) and by Bakhilina (1975: 47, 220–221). These expressions refer to the coats of horses,

cows or bullocks, with *buryj* indicating a colour between russet and (dark) brown. Later, by the 17th century, in folklore sources, fox and wolf coats are also named *buryj* (Slovar' Akademii Rossijskoj 1789–1794, vol. 1: 388; Bakhilina 1975: 67). Bakhilina (1975: 220–221) remarks that in Old Russian texts (e.g., 16th century) the use of *buryj* was not restricted to the animal domain: it collocated, e.g., with 'stone', 'hat', 'hair', 'eyes' and 'face' (see also Supplement S3).

Bakhilina (1975: 224, 227–228) deemed that in Old Russian and later, *buryj* appeared to have functioned as an abstract, basic term for 'brown'. Her conclusion is supported by the fact that in the Dal dictionary, one of the largest dictionaries of the Russian language, there was a detailed and exemplified entry for *buryj* (Dal 1863–1866, vol. 1: 144), while the entry for *koričnevij* appeared only under *kora* 'bark', with the term defined via *buryj* and as the colour of cinnamon (Dal 1863–1866, vol. 2: 160). In modern Russian, markedly, *buryj* collocates exclusively with 'bear', 'hare', 'coal', 'ore', 'wheat', 'soil', 'dust', 'muddy water', etc. Within the BROWN cluster, obtained from a psycholinguistic study by Frumkina (1984: 68), *buryj* is peripheral to the *koričnevij* "core" (see Supplement S4).

Koričnevij, in comparison, is recorded no earlier than the 17th century. It is a denominal adjective derived from Russian *корица* / *korica* 'cinnamon', the name of its referent object (Slovar' Akademii Rossijskoj 1789–1794, vol. 3: 821; Herne 1954; Bakhilina 1975; Slovari 11–17 vekov 1975–). The ubiquitous use of *korica* in Russian households by the 16th and 17th centuries (see Supplement S5) apparently bestowed on the spice the role of a cognitive "anchor" as the appropriately coloured object. This, in turn, prompted the linguistic development of the word *korica* – including the emergence of variations of its adjectival forms and meanings, as well as its essential role in naming the BROWN colour space area in modern Russian.

In modern Russian (printed texts) from the 1980s–1990s, linguistic measures and derived rankings of the two 'brown' terms were compared in studies from the University of Surrey (Corbett & Morgan 1988; Morgan & Corbett 1989; Davies & Corbett 1994; Corbett & Davies 1995). In a listing task, *koričnevij* ranked 9.5, i.e., within the rank range of BCTs, whereas *buryj* ranked 41.5 (Davies & Corbett 1994: 73); in comparison, based on the frequency-in-text scores, *koričnevij* ranked 10 and *buryj* 11; however, based on derivational productivity, *buryj* ranked 9 compared to *koričnevij* ranking 14 (Corbett & Morgan 1988: 57). By referring to Worth et al.'s (1970) and Hill's (1972) Russian derivational analyses, the

authors commented that *buryj*'s strong foothold could partly be explained through its use in specific collocations.

The corpus-based frequency dictionary of the Russian language published ca. 20 years later (Lyashevskaya & Sharov 2009) indicates a rank of 25 for *koričnevij* among colour terms, whereas *buryj* does not feature in the authors' ranking list at all. In a recent psycholinguistic study, *koričnevij* ranked 6, while *buryj* occurred very rarely, predominantly in the jocular vernacular expression *sero-buro-malinovij* (lit. 'grey-*buryj*-raspberry') used by Russian speakers to refer to a drab, nondescript colour shade (Paramei et al. 2018).

1.7 Hypothesis and aims of the present study

In the present study we aimed to ascertain, from a diachronic perspective, the linguistic mechanisms of colour term evolution. Specifically, we explored the linguistic behaviour of the two "Russian browns", *buryj* and *koričnevij* – i.e., the frequency of their occurrences and combinability with nouns signifying certain classes of objects. For this purpose, we undertook a computational analysis spanning more than two centuries of Russian books.

We tested Rakhilina's (2007a, 2007b, 2008) hypothesis that an old colour term applies to natural objects, whereas an incipient colour term initially collocates with nouns denoting artefacts but gradually expands to the realm of natural objects. In our analysis we leaned upon the promising results of our previous study, where the Russian National Corpus was employed to ascertain frequencies of collocations of the two "Russian browns" with nouns denoting natural objects and artefacts (Rakhilina & Paramei 2011).

2. Materials and Methods

2.1 Data source: Russian subcorpus of Google Books Ngram

We employed the Russian subcorpus of Google Books Ngram (GBN; Michel et al. 2011), which contains data on the frequencies of individual words, as well as *n*-grams, i.e., contiguous sequences of *n* words, with *n* = 2, 3, 4 or 5. It is the largest corpus of the Russian language, which makes it a valuable tool for studies of language evolution (Richey & Taylor 2020; Solovyev et al. 2020), although it is criticised by some as being unbalanced (Pechenick et al. 2015; Koplenig 2017). The second version of the GBN Russian subcorpus was released in 2012

(Lin et al. 2012). It included texts of 510,310 Russian books published between 1607–2009, with a total number of 67 billion words.

In the present study, the third version of the GBN Russian subcorpus was predominantly used (Google Books Ngram 2020), which has a significantly greater number of sources and covers an extended timescale: it includes the texts of 1,091,000 books published between 1486–2019, for a total number of 89 billion words. In the third version, greater attention was paid to the quality of the text recognition; errors in the dating of the books were corrected, and some erroneously included books were excluded. (As a result, in spite of the total growth of the corpus size, the number of books from certain years decreased.) Furthermore, the third version only includes 2-grams (bigrams) that occurred at least 40 times in the corpus for the entire recorded period. For this reason, and due to the corpus corrections, some *n*-grams recorded in the second corpus version were omitted in the third version. To overcome these omissions, for our analysis we selected bigrams using both the second and third versions of the corpus.

Notably, the majority of the books contained in the Russian subcorpus of GBN were published after the beginning of the 19th century. Since the data for the distributional analysis of *buryj* and *koričnevij* only become sufficient and representative from 1820, in the present study we focussed on the period comprising the last 200 years.

2.2 Data cleansing and lemmatisation

For the analysis, we extracted the frequencies of all bigrams corresponding to attributive constructions with *buryj* and *koričnevij* (including their inflectional forms). Along with the basic forms of these CTs, we also considered their prefixed (e.g., *zaburet'* 'to become *buryj*') and suffixed derivatives (e.g., *koričnevatyj* 'with a *koričnevij* shade'). Further, we recorded compounds (e.g., *sero-buryj*, *sero-koričnevij* 'grey-brown') and forms with achromatic modifiers (e.g., *svetlyj* 'light'; *tëmnyj* 'dark'; *blëklyj* 'pallid' etc.). Since in early texts the spelling of the incipient 'brown' term was unstable, we also recorded, along with the present form *koričnevij*, collocations with its variants *korišnyj*, *korišnevij* and *koričnyj*. Bigrams of the types *Noun+Colour term* and *Colour term+Noun* were selected automatically.

Of note is the fact that GBN, a part-of-speech (POS) tagged corpus, contains numerous POS-tagging errors. To rectify inaccuracies, in order to lemmatise the nouns that collocate with the terms *buryj* and/or *koričnevij*, the POS-tagged data were verified using the

OpenCorpora morphological dictionary (OC; OpenCorpora n.d.), one of the largest electronic dictionaries of the Russian language, which currently contains 391,800 lemmas that include 5,140,000 word forms (see also Bocharov et al. 2013). Finally, to ensure that only the target bigrams were selected, in some cases, a manual check was also performed.

In total, 3,966 bigrams were selected, including words relating to 1,116 different lemmas. The bigrams were divided into several classes following two non-exclusive approaches. One approach distinguished the nature of the object denoted by the noun with which the Russian ‘brown’ terms collocate, which resulted in the following four classes of objects, or “nominal fields”, according to Steinvall (2002):

- *natural objects*
- *artefacts*
- *abstract concepts*
- *figurative senses*.

Note that some Russian nouns collocating with ‘brown’ terms may fall into more than one class depending on the context. For instance, the (diminutive) noun *peněk* (derived from *pen’* ‘stump’) has the meaning of either a ‘tree stump’ or a ‘tooth stump’ (metaphor); hence, bigrams with this noun were classified as either “natural objects” or “figurative senses” respectively. Similarly, the (diminutive) noun *spinka* (derived from *spina* ‘back, spine’) can denote either the back of a pet or of an (arm)chair, with the corresponding bigrams falling into either “natural objects” or “artefacts”.

The other analytical approach was focussed on distinctive collocations of the ‘brown’ terms with specific nouns or certain noun classes. In particular, we discerned bigrams that collocate solely with *buryj*, solely with *koričnevij*, or with either term. Examples of *buryj*-only collocations are *lisa* ‘fox’ or *boloto* ‘marsh’, of *koričnevij*-only collocations is *platje* ‘dress’, while the nouns *voda* ‘water’ and *sneg* ‘snow’ are recorded with either term.

Selection of the bigram lists and their lemmatisation, extraction of bigram frequencies from the GBN Russian subcorpora (second and third versions), and statistical analysis were performed using scripts written in the Matlab environment.

3. Results

3.1 Dynamics of the occurrences of *buryj* and *koričnevij*

As indicated in §2.2, 1,117 nouns were identified which collocate with the terms *buryj* and *koričnevij*. From these, 156 nouns co-occur only with *buryj*, 577 only with *koričnevij*, and 384 appear in combinations with either term. Thus, over the recorded period (1486–2019), in general, *koričnevij* collocates with more nouns than *buryj*.

Figure 1 shows the dynamics of the frequency of the terms *buryj* and *koričnevij* between 1800–2019 (when data are sufficient for distributional analysis of the terms). It is apparent that at the beginning of the 19th century, the frequency of *buryj* was significantly greater than that of *koričnevij*. In the following century, however, *buryj*-frequency steadily decreased, while *koričnevij*-frequency incrementally but steadily increased, especially in the last 100 years. Finally, from the beginning of the 1980s, after a long period of competition with *buryj*, collocations with *koričnevij* started to prevail.

It is worth bearing in mind that in the Russian GBN subcorpus, some objects are mentioned quite often in combination with the ‘brown’ terms, whereas other objects are mentioned only several times per century. Hence, the observed dependencies can be due to co-occurrences of the two ‘brown’ terms with a relatively small number of frequently used nouns, which might disguise the ongoing competition of the “Russian browns” in typical co-occurrence cases.

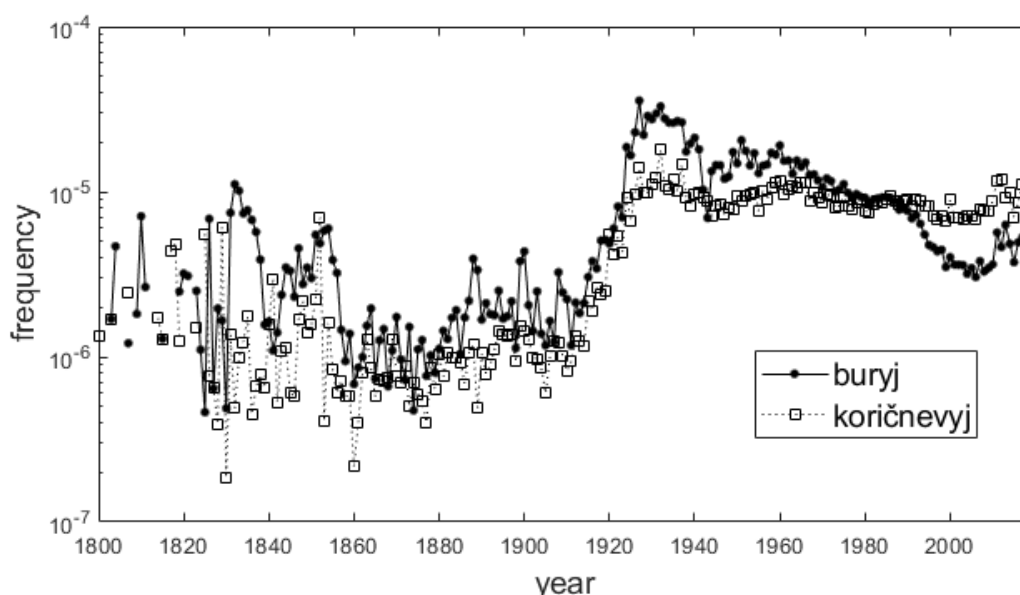


Figure 1. Frequency of occurrences of *buryj* and *koričnevij* in the Russian subcorpus of Google Books Ngram (1800–2019)

3.2 Derivational morphology of *buryj* and *koričnevij*

The GBN records attest to a comparable derivational morphology for the two ‘brown’ adjectives; we focus on the modified and compounded adjectives, and the verbal and nominal derivatives of the terms (cf. Corbett & Morgan 1988). We observe that the variety of specific modifiers differs between the two terms. As indicated in Supplement Table S1, modern Russian of the beginning of the 21st century possesses a much greater variety of derivational forms of the two ‘brown’ terms than those indicated by Corbett and Morgan (1988) for the 1950s–1970s.

In particular, *koričnevij* acquired a novel modification *tëmno-* ‘dark’, and this overtook the previously recorded modifier *svetlo-* ‘light’ (Figure 2). Also, compared to *buryj*, *koričnevij* is used more frequently with achromatic modifiers and in a compound with *seryj* ‘grey’ (see Supplements Figures S1–S3). Conversely, *buryj* gained a new modifier *svetlo-* ‘light’, along with previously recorded *tëmno-* ‘dark’ (Supplement Table S1). Further, on the Internet (only) both “Russian browns” appear with the modifier *blëklo-* ‘pallid’, not reported earlier (Corbett & Morgan 1988), but these instances are very rare and found in texts no earlier than 2014 (see Supplement SE1).

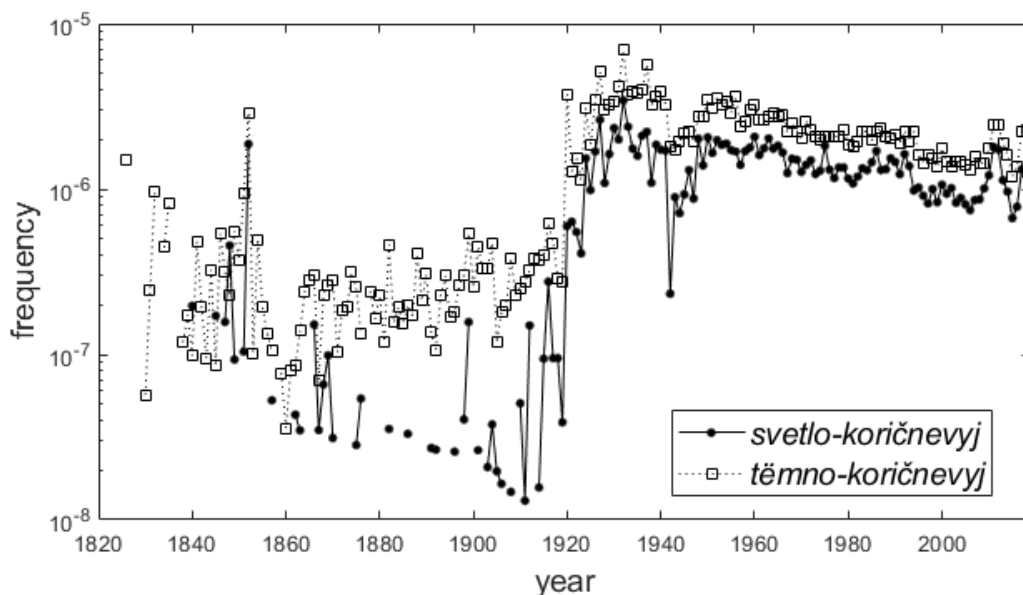


Figure 2. Diachronic dynamics of *koričnevij* with modifiers *tëmno-* ‘dark’ (.....) (темно-коричневый / темнокоричневый) and *svetlo-* ‘light’ (—) (светло-коричневый / светлокори́чевый) in the Russian subcorpus of GBN (2019)

In contrast, the variety and frequency of the verbal derivatives of *buryj* are several orders greater than those of *koričnevij*. Figure 3 illustrates the dynamics of the frequently used verb *buret'* 'to acquire *buryj* colour' (cf. English *to redden*). Also recorded are incentive (prefixed) *buryj*-verbal forms ($\emptyset\sim et'$ '-en') *poburet'* 'to become *buryj*' and *zaburet'* 'to begin acquiring *buryj* shade' (see further examples in Supplement SE2). In comparison, inceptive verbal forms *pokoričnevet'* and *zakoričnevet'* 'to become *koričnevij*' are very rare, and can be found only on the Internet or in a recent modern Russian dictionary (Lopatin 2013). Note also that in the last ca. 70 years the usage of *buryj* verbal derivatives has been falling steadily, whereas the frequency of *koričnevij*-derived verbal forms has been rising, albeit quite modestly.

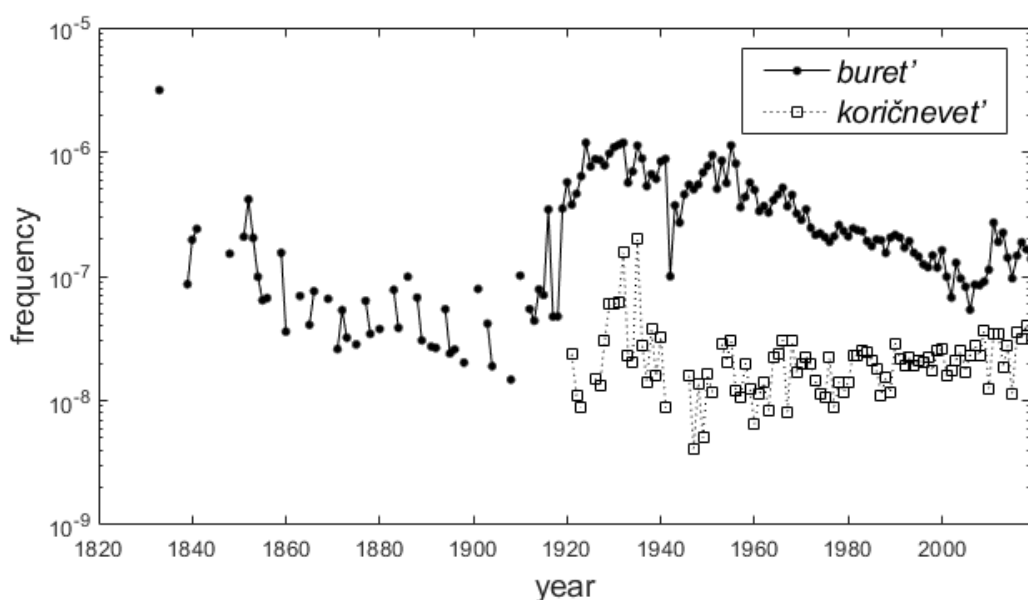


Figure 3. Diachronic dynamics of verbal forms $\emptyset\sim et'$ '-en' derived from *buryj* (—) and *koričnevij* (- - -). Note that low *koričnevij* frequency is multiplied by 5 for a better visualisation

Recently, although rarely, instances of the nominal derivatives *-izna* '-ness' of both 'brown' terms have been recorded: *burizna* has been attested in GBN and on the Internet, while *koričnevizna* has been recorded solely on the Internet (see examples in Supplement SE4) or in specialist publications on the Russian language (Ulukhanov 2015).

To summarise, across the derivatives of the two 'brown' terms, we observe a transitional state: *buryj* still retains rich derivational productivity indicative of a BCT; in parallel, a greater variety of derivational forms is emerging for *koričnevij*.

3.3 Frequency distribution of “Russian browns” in diachrony

We scrutinised the diachronic dynamics of the diversity of usage of the two Russian ‘brown’ terms between 1820 and 2019. This was achieved by calculating the number of different nouns that collocate with either *buryj* or *koričnevij* within consecutive time intervals. To obtain a smooth function, 5-year intervals were considered for the calculations (from 1820–1824, 1825–1829, ..., 2015–2019). Results are shown in Figure 4A.

Note that some of the identified collocations are quite frequent in the corpus while others occur only once in several years. To factor in the unevenness of bigram occurrences, we calculated entropy of the frequency distribution, proposed for processing corpus data (Juola 2003), as a measure of lexical diversity. In particular, we computed the information entropy (h) of the frequency distribution of each ‘brown’ term in bigrams with various nouns. Expressed in bits, the entropy measure is, however, not particularly telling. Therefore, we computed a more instructive entropy derivative – perplexity of frequency distribution, 2^h (Brown et al. 1992):

$$\text{Perplexity} = 2^h, \quad h = - \sum_i p_i \log_2 p_i$$

where p_i is the relative frequency of various nouns in bigrams with each of the ‘brown’ terms. Perplexity reflects the number of frequently used noun alternatives, i.e., those collocating with either *buryj* or *koričnevij* (see Figure 4B).

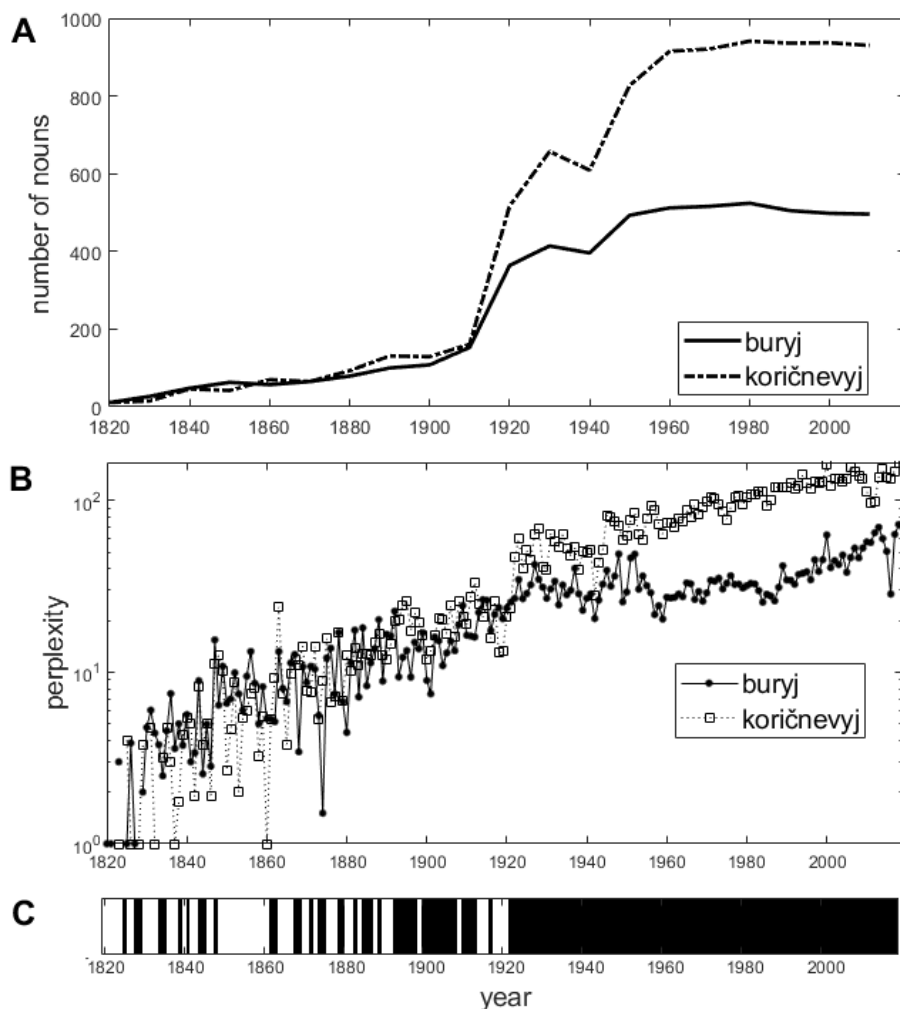


Figure 4. (A) Frequency distribution of bigrams containing a noun with either *buryj* or *koričnevij*, computed year-by-year. (B) Perplexity of frequency distribution of bigrams containing a noun with either *buryj* or *koričnevij*, computed year-by-year. (C) The prevalent term perplexity for *buryj* (white) and *koričnevij* (dark brown)

Figure 4A, B demonstrates that the number of nouns combining with either *buryj* or *koričnevij* increases after 1820 – primarily due to the growth of the corpus size, as a manifestation of Heaps’ law (Heaps 1978; van Leijenhorst & van der Weide 2005). It is also apparent that initially, more nouns collocate with *buryj* than with *koričnevij*; however, after a period in which the two terms compete, from the 1920s onwards, *koričnevij* starts to prevail. Furthermore, the post-WWII diversity of the denoted objects combined with *koričnevij* becomes perceptibly greater compared to those combined with *buryj*.

In an alternative form, the dynamics of the prevalence of perplexity for the two “Russian browns” is presented in Figure 4C. It reveals an initially greater collocational

diversity of *buryj* (until ca. 1900), then a period of competition between the two terms (around 1900–1920), followed by the increasing importance, in the 1920s, of *koričnevijj*, and its further incremental rise from the mid–1940s onwards.

3.4 Competition of the two Russian ‘brown’ terms

As indicated in §3.1, we were able to identify 384 nouns that combine with either *buryj* or *koričnevijj* (*b/k*-nouns). To examine the terms’ competition in the period under consideration, for each year, we estimated the frequency of collocation of each of the two terms with the *b/k*-nouns. Specifically, for each year, the percentage of *koričnevijj*-collocation instances was calculated for each of the nouns; the values of *koričnevijj*-percentage were then averaged over the 384 nouns. Figure 5 presents the average percentages of *koričnevijj* in the double-collocations over the last 200 years. Note that the mean may not always be indicative, since a small number of strong bursts (for individual words) could significantly distort the mean value; therefore, median values were also calculated.

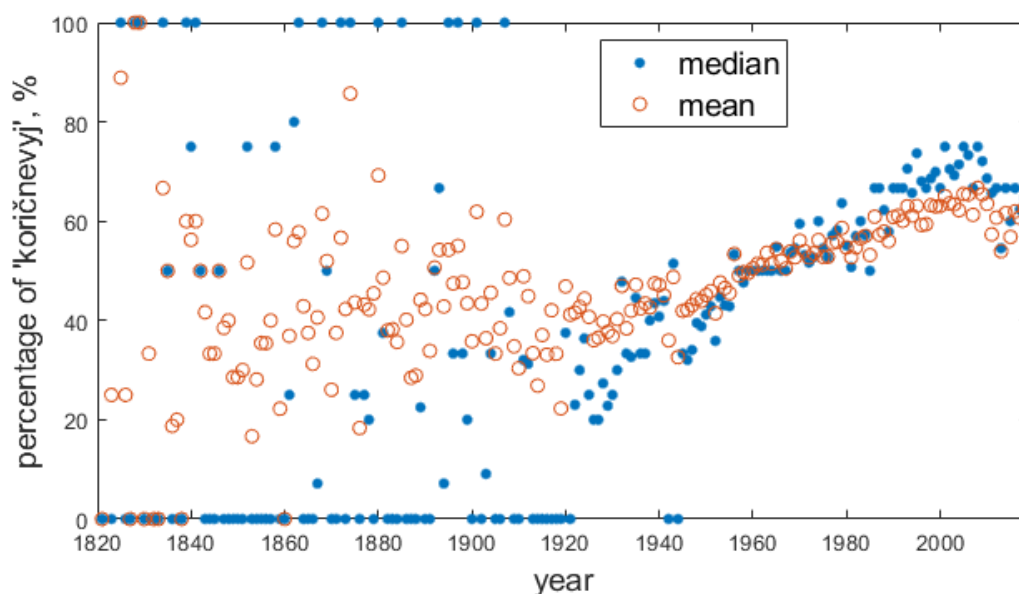


Figure 5. Dynamics of the average percentages of *koričnevijj*-collocations, from the total number of uses with either *buryj* or *koričnevijj* (*b/k*-nouns), in the sample of 384 nouns in the GBN Russian subcorpus (1820–2019)

As shown in Figure 5, the percentage of *koričnevijj* use varies greatly from year to year before ca. 1920, probably due to the small size of the corpus in those years. The function then

becomes less noisy with a much greater corpus size due to a huge increase in production and dissemination of (mainly new) books after the October Revolution of 1917. We compared the average percentages of *koričnevyyj*-collocations for each of the 384 nouns for two periods “flanking” the 20th century – one before the October Revolution, 1850–1917, and the other one being the most recent, 2000–2019. We found that for the pre-1917 period, the average *koričnevyyj*-share was 29%, compared to a much higher 62% share for the last 20 years, attesting to an increase in *koričnevyyj* use during the 20th century.

Note that the increase in the (average) percentage of *koričnevyyj* use does not imply that this was the case for any of the *b/k*-nouns. To take a closer look at individual nouns, we compared their linguistic behaviour in the first and second half of the 20th century. Specifically, we calculated the total frequencies of all bigrams in two intervals, 1900–1945 and 1950–2019, and estimated the percentage of *koričnevyyj* combinations with each noun in each of the two intervals. Figure 6 shows a scatterplot of the percentage of *koričnevyyj* use in 1900–1945 vs. 1950–2019. It is apparent that the share of *koričnevyyj* use increased for most double-collocating nouns, estimated as 74.5%, but also decreased for some.

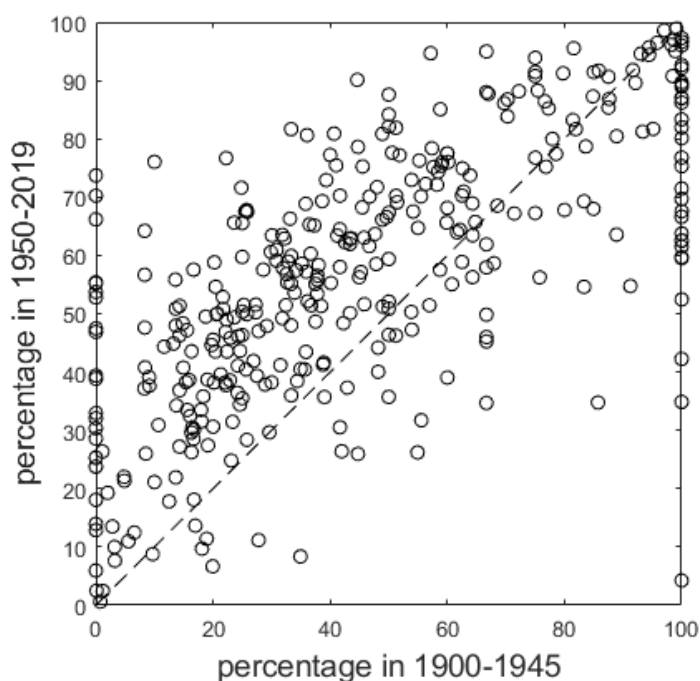


Figure 6. Scatterplot of the percentage of *koričnevyyj* use in relation to the total number of uses with the 384 nouns collocating with both *buryj* and *koričnevyyj* (*b/k*-nouns), in 1900–1945 and 1950–2019. Note that each dot represents an individual noun

We further scrutinised the dynamics of the percentage of *koričevyj* in *b/k*-nouns for each year between 1920 and 2019. For this, the kernel density estimation was obtained using a 5.9% half-width triangular window. Since the function remained quite noisy, it was further smoothed using a moving average with a 3-year half-width triangular window.

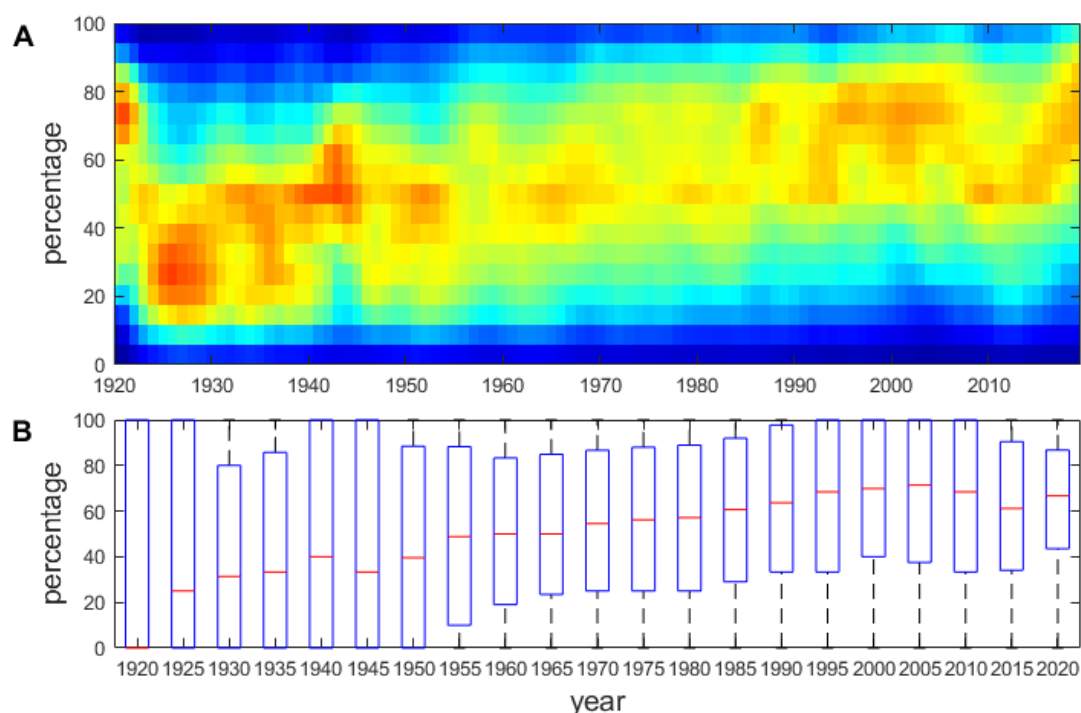


Figure 7. Distribution of the average percentage of *koričevyj* use in the period 1920–2019. (A) Distribution density (heatmaps) for each year. (B) Distribution variation (boxplots) for 5-year intervals

A heatmap of the distribution density function (Figure 7A) demonstrates that the percentage of *koričevyj* in *b/k*-nouns has been steadily growing over the last 100 years. This observation is confirmed by a complementary Figure 7B illustrating the dynamics of the distribution variation of the *koričevyj* percentage by medians in 5-year intervals.

3.5 Nouns of object classes in bigrams: A diachronic analysis

We classified 1,116 nouns identified in the GBN corpus according to the following semantic classes:

- **natural objects:** physical objects in nature that have a dominant diagnostic colour (e.g., *babochka* ‘butterfly’, *kamen’* ‘stone’, *pustynya* ‘desert’, *more* ‘sea’, etc.)

- **artefacts:** man-made objects (e.g., *avtomobil* ‘car’, *botinok* ‘boot’, *gorshok* ‘pot’, *fetr* ‘felt’, etc.) that can be coloured in various artificially-produced colours
- **abstract concepts:** concepts which do not possess a physical form, size or colour (e.g., *prostranstvo* ‘space’, *kolorit* ‘complexion/cast’, *svet* ‘light’, *t’ma* ‘gloom’, etc.)
- **figurative senses** (metonymies and metaphors): cases in which the ‘brown’ terms do not denote a particular colour (e.g., *koričnevaya čuma* ‘brown plague’, *koričnevaya zaraza* ‘brown pest’, *koričnevaya diktatura* ‘brown dictatorship’, etc.).

All the nouns were further divided into three groups depending on how they collocate with *buryj* and *koričnevyyj*:

- only with *buryj* (156 nouns)
- only with *koričnevyyj* (576 nouns)
- with both *buryj* and *koričnevyyj* (384 nouns).

For each of the four classes of nouns, Table 1 shows the number of nouns belonging to each of the three groups of the ‘brown’ terms’ collocations. Note that the sum of the numbers in the first four lines is not equal to the sums in the ‘Total’ boxes, since many nouns are polysemous, i.e., they belong to more than one class. Figure 8 illustrates the distribution of the terms’ collocations across the four semantic classes.

Table 1. Number of the nouns (including polysemous ones), within each of the four semantic classes, collocating either solely with *buryj*, or solely with *koričnevyyj*, or with both *buryj* and *koričnevyyj*

Class of nouns	Only with <i>buryj</i>	Only with <i>koričnevyyj</i>	With both <i>buryj</i> and <i>koričnevyyj</i>	Total
Natural objects	142	176	311	629
Artefacts	17	413	132	562
Abstract concepts	1	10	15	26
Figurative senses	0	31	4	35
Total	156	576	384	1,116

Table 1 and Figure 8 make it apparent that among the nouns that collocate solely with *buryj*, 91.0% denote natural objects. In contrast, in collocations solely with *koričnevyyj*, nouns denoting artefacts prevail (71.7%), while a mere 30.6% of nouns denote natural objects. We also found four cases, in which nouns denoting natural objects collocate with both *buryj* and

koričnevijj; however, in all cases of *koričnevijj*-collocations the noun is used figuratively: *zver* ‘beast’, *zmeja* ‘snake’, *krov* ‘blood’, *plesen* ‘mildew’ (see “Figurative senses” in Table 1). Of note, of the total 35 cases of figurative senses, all involve *koričnevijj*-collocations and are associated with fascism.

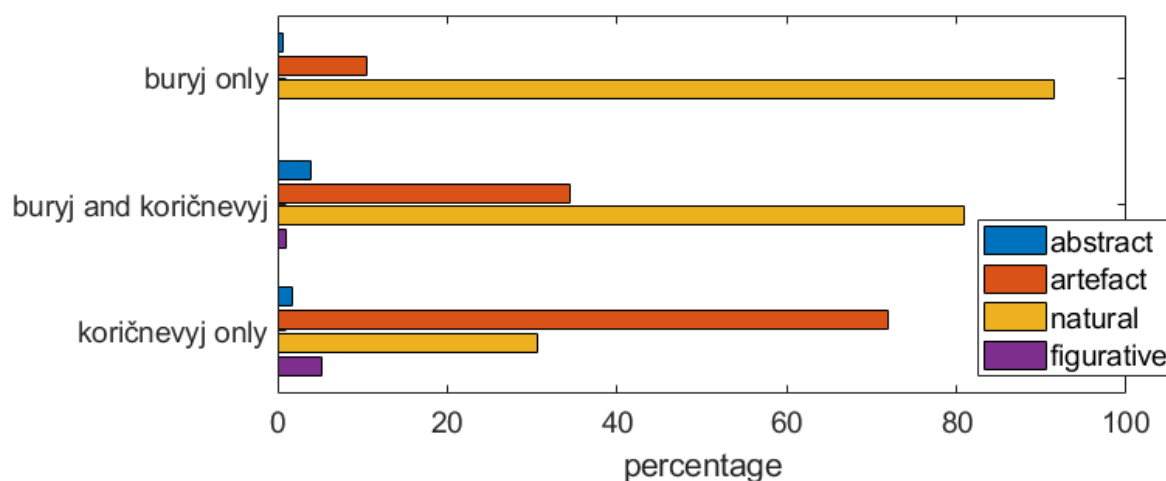


Figure 8. Percentage of nouns of the four semantic classes (including polysemous ones) that collocate solely with *buryj*, solely with *koričnevijj*, or with both *buryj* and *koričnevijj* (*b/k*-nouns)

Let us scrutinise the dynamics of the ‘brown’ terms with the 384 double-collocating nouns. For this, the percentage of *koričnevijj*-collocations was calculated with nouns of the following six categories:

- Nouns denoting only
 - natural objects (233)
 - artefacts (61)
 - abstract concepts (12)
- Polysemous nouns that denote
 - either natural objects or artefacts (71)
 - either natural objects or abstract concepts (3)
 - natural objects but can be used in a figurative sense (4)

In the last two categories, there was a paucity of occurrences for statistical interpretation; hence, in the following, we present results for the first four categories of nouns only.

Next, for each of the four noun categories we calculated the median percentage of *koričnevij*-collocations. Results are shown in Figure 9. Note that it presents data for the last hundred years, from 1920 onwards, with many more books in the GBN corpus and, hence, more reliable values, compared with a considerable dispersion of values for the period before 1920.

Figure 9 shows that the percentage of *koričnevij* collocations progressively increases in all four noun categories. However, a visual inspection of function slopes hardly allows inter-category comparison of the rate of transition from *buryj* to *koričnevij*. Apparent, though, is the difference between individual categories in the time-point of such a transition: front runners in collocating with *koričnevij* are nouns for artefacts, while nouns for natural objects lag behind; in-between is the function for polysemous nouns, denoting both natural objects and artefacts. Similar behaviour is observed for nouns of abstract concepts, which “become” *koričnevij* earlier than natural objects, although the function has a large dispersion due to a small number of nouns in this category.

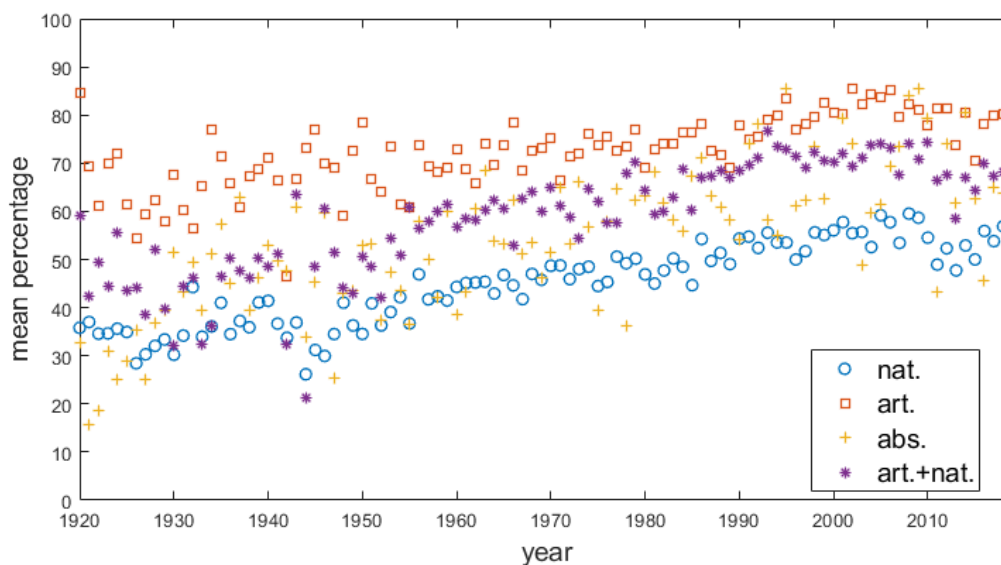


Figure 9. Change of the mean percentage of *koričnevij*-collocations, out of the total number of collocations containing either ‘brown’ term, for each of the four noun categories: nat. – natural objects (233 cases); art. – artefacts (61 cases); abs. – abstract concepts (12 cases); art.+nat. – polysemous, either artefacts or natural objects (71 cases)

Thus, an analysis of collocations in the GBN database provides evidence of the transition from *buryj* to *koričnevij*. One observes, however, considerable variation among

individual nouns, including instances of an increase in percentage of *buryj*-collocations. To estimate the rate of transition to *koričnevij* in each of the four noun categories, we undertook a linear approximation of the transition-to-*koričnevij* function (%) for the interval between 1950 and 2005. We applied the method of iteratively reweighted least squares, a robust algorithm that mitigates the influence of outliers in an otherwise normally-distributed data set (Holland & Welsch 1977; Street et al. 1988). Specifically, we estimated a linear regression coefficient characterising the rate of transition-to-*koričnevij*- for each noun. A positive coefficient value indicates an increase in the percentage of *koričnevij*-collocations; conversely, a negative value indicates an increase in the percentage of *buryj*-collocations for the noun in question. The greater the modulus of the linear regression coefficient, the faster the occurring transition.

Figure 10 shows boxplots of regression coefficients characterising the rate of transition from *buryj* to *koričnevij* for the four noun categories. As is apparent, median values of the transition rate are comparable for the four categories, however, the regression coefficients vary significantly within categories.

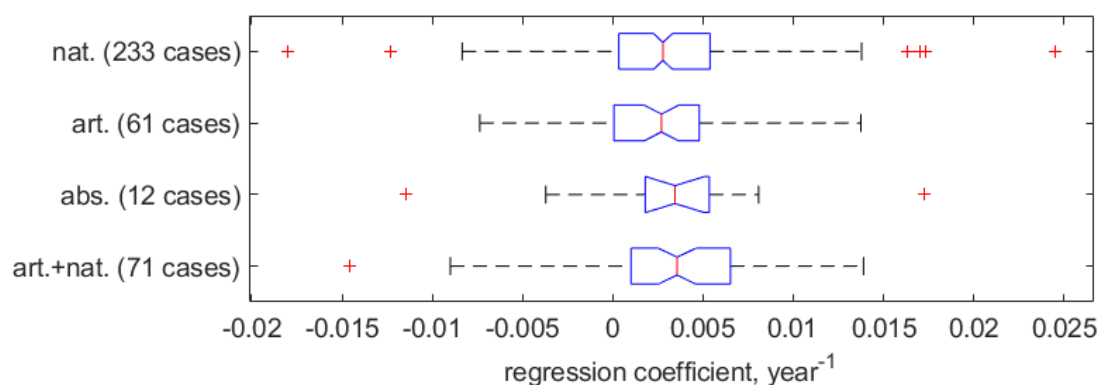


Figure 10. Boxplots of regression coefficients characterising the rate of transition (%) from *buryj* to *koričnevij* for the four noun categories: nat. – natural objects; art.- artefacts; abs. – abstract concepts; art.+ nat. – polysemous, either artefacts or natural objects

We assessed inter-category differences of the transition-to-*koričnevij*- rate using the Wilcoxon rank-sum test (Gibbons & Chakraborti 2010), whereby linear regression coefficients for the category nouns were compared with those for the nouns in the other three categories (cumulatively). For the four comparisons, the following *p*-values were obtained: 0.4999, 0.0839, 0.1133, 0.3082, respectively (in the same order as in Figures 9 and 10), which implies

no significant inter-category differences in the transition rate. A similar outcome was obtained when we performed pairwise comparisons between the noun categories (Table 2), with $p=0.1118$ being the lowest for differences between collocations with the nouns denoting artefacts and polysemous nouns.

Table 2. Pairwise comparisons of the rate of transition from collocations involving *buryj*- to those involving *koričnevij*- between the four noun categories; indicated are p -values

Noun category	Artefacts	Abstract concepts	Polysemy (natural objects or artefacts)
Natural objects	0.5816	0.6147	0.1349
Artefacts	-	0.4189	0.1118
Abstract concepts	-	-	0.8697

Thus, there are no significant differences in the transition-to-*koričnevij* rate between the noun categories. For natural objects and abstract concepts, the transition unfolds at the same rate as for artefacts but occurs with a lag. One can estimate the characteristic time interval of transition (CTIT) from *buryj* to *koričnevij*. Let us define the CTIT as the time interval (in years), during which the percentage of *koričnevij* will increase by 50% (for example, from 30% to 80%). We found that for the entire sample of double-collocating 384 nouns, the median of the linear regression coefficient corresponding to the CTIT is 175.8 years.

3.6. Analysis of atypical contexts where *koričnevij* collocates with nouns for natural objects

The results presented in §3.1–3.5 support Rakhilina’s (2007a, 2007b, 2008) hypothesis that *buryj*, the old term, collocates with nouns denoting natural objects, whereas more recent *koričnevij* initially applies to artefacts but gradually expands to the realm of natural objects. Analysis of individual examples indicates, however, that some instances violate this collocation pattern, whereby nouns for natural objects collocate predominantly or solely with *koričnevij*. We scrutinised such atypical cases to fathom the linguistic motivation of *koričnevij* use and appraise the context types.

Firstly, we more closely examined *koričnevij*-collocations with nouns for animate natural objects, exemplified by *medved* ‘bear’/*medvežonok* ‘bear cub’, *zmeja* ‘snake’, and

pauk ‘spider’ (see Supplements SE5–SE13). Based on an overview of the recorded cases, we can identify the following contexts:

- Russian translations of ‘brown’ in fiction from other languages
- Russian translations of zoological terms from English
- References to a fairy-tale character, in native texts and Russian translations
- In native Russian fiction, as part of a metaphor (e.g. the snake-like shape of a scarf), where an artefact alludes to an animal.

Cases of *medved’* ‘bear’, the noun deeply entrenched in Russian in combination with *buryj* (see Figure 11), reveals two further contextual types of collocations with *koričnevij*:

- descriptions of a bear in paintings and appliqués, in children’s drawings, and (a black bear) as one of the national symbols of Canada
- in references to toys (teddy bears) and souvenirs (as the diminutive *medvežonok* ‘bear cub’).

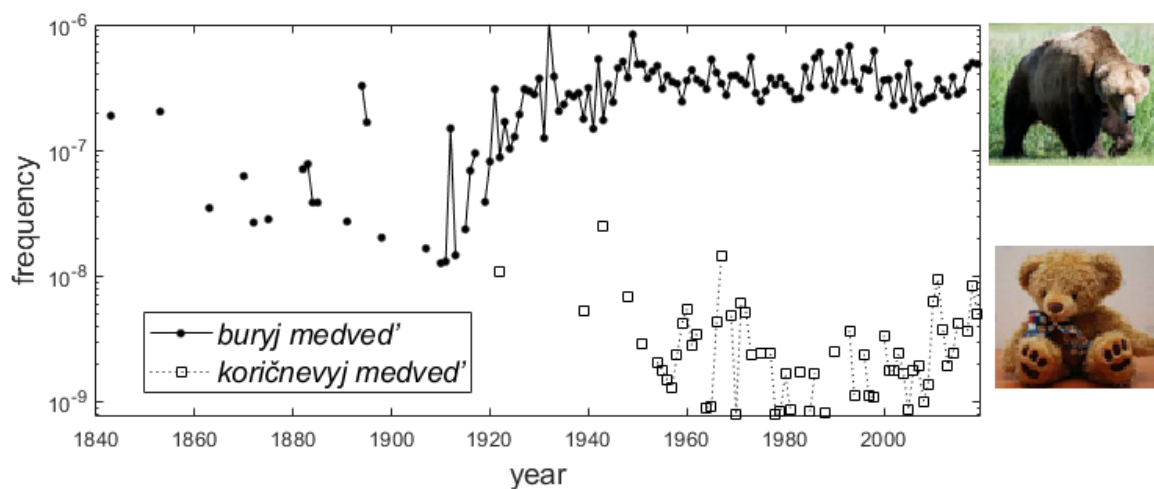


Figure 11. Frequency of occurrences of *buryj medved’* and *koričnevij medved’* in the Russian subcorpus of Google Books Ngram (1840–2019)

Note that in the cases of bear representation, synthetic pigments and dyes are used to render the artefact’s colour, resulting in a homogeneous and vivid brown colour.

Second, *koričnevij* collocates with nouns depicting minerals, i.e., natural inanimate objects, that have been processed to become artefacts, such as *mramor* ‘marble’ used for flooring (Supplement SE14), or *almaz* ‘diamond’ embedded in an ornament (Supplement

SE15). In many such cases the Russian colour term is a translation of *brown* from English-language advertisements.

Third, *koričnevij* collocates with nouns depicting some other inanimate natural objects or substances, such as human body parts (e.g., *lik* ‘face, countenance’), *bryzgi* ‘splashes’, *nagar* ‘soot’, *gaz* ‘gas’. In two cases of food products, *ris* ‘rice’ and *saxar* ‘sugar’, *koričnevij* implies unrefined rice or sugar respectively. Usually it is the translation of English *brown* in the name of an imported (as a rule, quite expensive) brand (Supplement SE16). It is worth noting that since these food products are of light brown colour with grey admixture, occasionally in the product labelling of rice and sugar, *koričnevij* is accompanied by an elucidative *buryj*, to ensure Russian consumers’ comprehension.

The fourth contextual type is that of literary images in fiction and essayistic writings, in which *koričnevij* is part of political vocabulary metonymies (cf. *nečist* ‘scum’, *armija* ‘army’, *čuma* ‘plague’; Supplements SE17–SE20) and allude to the colour of *Braunhemden* ‘Brownshirts’, a paramilitary wing of the German Nazi party.

4. Discussion

4.1 Summary of the major findings

In the present study, we explored the linguistic mechanisms of colour term evolution, specifically the terms’ contextualised linguistic behaviour, and compared this to currently prevailing decontextualised denotation-based methodology. We undertook a diachronic computational analysis of the two competing Russian terms for ‘brown’ – the old *buryj*, glossed as ‘dust/greyish brown’ (11th century), and the historically new *koričnevij* (17th century), a basic ‘brown’ term in modern Russian – to investigate the dynamics of their distributional semantics. The study was motivated by Rakhilina’s (2007a, 2007b, 2008) hypothesis of linguistic colour term evolution: that an old colour term applies to natural objects, whereas an incipient colour term, its contestator, initially applies solely to artefacts but gradually expands to the realm of natural objects.

For this purpose, we used a Russian subcorpus of the third version of Google Books Ngram (2020) to ascertain factors underlying the process of supplanting the old term by the new challenger, as well as the timescale of establishing *koričnevij* and the pace of its

entrenchment in Russian. By applying a time-series analysis we compared the combinability of the two terms in bigrams with nouns that signify various object categories: natural objects, artefacts, abstract concepts, and figurative senses. We were particularly interested in the (relative) frequencies of occurrences of *buryj* and *koričnevyyj* in bigrams with nouns (N=384) collocating with either of the ‘brown’ terms in a representative book sample from the last 200 years, 1800–2019.

The results provide evidence that in total frequency of use, *buryj* was initially dominant and, in its collocations, remained virtually unchanged across centuries; however, it started to go into decline during the 19th century. Concurrently, its ‘contester’ *koričnevyyj* became increasingly used for brown denotation (see Figure 1): *koričnevyyj* overtook *buryj* at the beginning of the 1920s and progressively prevailed from the beginning of the 1960s. Furthermore, the perplexity index indicates a significant increase in the scope of nouns collocating with *koričnevyyj*, initially at the beginning of the 1920s and with another upsurge in the mid-1940s (see Figure 4). The findings on the expansion of the *koričnevyyj* collocational potential are complemented by the gradual increase of the Jensen-Shannon divergence between the frequency distributions of *buryj* and *koričnevyyj* in the same two periods, as reported by us elsewhere (Bochkarev et al. 2021). These estimates of distributional semantics corroborate the status of *koričnevyyj* as the basic CT for ‘brown’ in modern Russian.

Notably, the two periods of upsurge of *koričnevyyj* use and distributional semantics are observed after the October Revolution in Russia (1917) and WWII respectively, two political upheavals reflected in the semantic field of colour (cf. Biggam 2006). We assume that the post-1920 upsurge in *koričnevyyj* use can be attributed to dramatic changes in Russian lifestyle and technology: in the new political era, obligatory secondary school education was introduced; occupational trends changed; the number of published books sharply increased; also, with the emergence of aniline pigments, the dyeing industry further developed, whereby the purchasers of clothes came to refer to brown-dyed cloths of all shades as *koričnevyyj*. The post-WWII term expansion apparently reflects its use in political vocabulary (see Supplements SE17–SE20).

The analysis of the combinatorial behaviour of the “Russian browns” for specific classes of nouns confirmed our earlier finding, based on the RNC analysis (Rakhilina & Paramei 2011) of the two terms’ division between the nominal fields, with *buryj* predominantly

collocating with nouns signifying natural objects and *koričnevij* initially collocating with nouns for artefacts but gradually expanding to natural objects.

Interestingly, scrutiny of atypical contexts, where nouns denoting natural objects (animals and minerals) collocate with *koričnevij*, reveals a twofold motivation for the majority of such cases: (i) translations into Russian from other languages, predominantly of English *brown*, using a modern Russian counterpart of the basic ‘brown’ term, and (ii) descriptions of pictorial representations of animals (e.g., a drawing of a bear; a teddy bear), or a processed state of minerals (e.g., marble flooring; an ornament-embedded diamond) (see Supplements SE5–SE16). The “artefact mode” of natural objects implied in (ii) appears to justify the lexical switching to the corresponding ‘brown’ term.

Importantly, in the present study, using the large-scale GBN corpus and methods of computational linguistics, we refined the analysis to explore, in addition, the use of the ‘brown’ terms in abstract concepts and figurative senses. We found a steady increase of *koričnevij*-collocations also in these nominal fields (see Table 1 and Figure 8).

Furthermore, the dynamics of distributional semantics, gauged for the *b/k*-bigrams, points out that during the last 200 years *buryj* has been increasingly replaced by *koričnevij* in the four examined noun categories – natural objects, artefacts, abstract concepts and polysemous nouns (denoting either natural objects or artefacts). We found that the rate of *koričnevij*-expansion is comparable across all categories (see Figure 10). However, in absolute numbers of the signified objects, the *koričnevij*-surge differs significantly between the categories, with artefacts being the front runners and natural objects being the last to “succumb” to the term replacement, with the other two categories being in-between (see Figure 9).

Together, the findings provide overwhelmingly conclusive evidence in support of Rakhilina’s (2007a, 2007b, 2008) hypothesis that an incipient colour term, *koričnevij*, establishes itself as a basic colour term gradually, by expanding into the realm of nouns signifying objects with a colour previously named by the old term, *buryj*. Moreover, the outcomes of the diachronic corpus analysis offer novel insights into the linguistic evolution of an emergent basic colour term – by revealing the process and pace of the new term’s increase in usage, and the expansion in its distributional semantics in various nominal fields, where it supplants the old term.

4.2 Denotata of *koričnevij* and *buryj*

Direct comparison of the denotative meanings of *buryj* and *koričnevij*, to our knowledge, has not been undertaken in a dedicated psycholinguistic study. However, the scope of collocations in the GBN bigrams revealed that the two terms differ with regard to the perceptual characteristics of the implied colour: *koričnevij* is used to allude to “pure”, homogeneous and vivid, saturated colour, whereas *buryj* is perceived as denoting “impure”, “diluted”, nonhomogeneous colour, with admixtures of grey or of shades of hues. In the GBN, the expression *rovnyj koričnevij* ‘equable/uniform *koričnevij*’ is recorded but not *rovnyj buryj*. Furthermore, Google returns the expression *čistyj koričnevij* ‘pure *koričnevij*’ 1,550 times but *čistyj buryj* only five times; similarly, the expression *odnotonnyj koričnevij* ‘monochromatic *koričnevij*’ produces more than 51,000 returns, compared with 700 returns for *odnotonnyj buryj*.

The origin of this perceptually-stipulated semantic dichotomy can be understood, if one keeps in mind that natural objects described in Russian as *buryj* – a bear, drab autumn leaves, dull-coloured stones, etc. – rarely have a pure, homogeneous colour with a well-defined hue; rather, their colour is a combination of different shades of brown, with admixtures of other colours – grey, ginger, etc. In comparison, since the mid-19th century, when synthetic aniline dyes were developed and underwent growing industrialisation, they were mostly used for dyeing artefacts (fabric, clothes, handbags, toys, etc.), becoming enablers of vivid, long-lasting, monochromatic colours (cf. Nicklas 2017). In the GBN, the collocation *koričnevij kostjum* ‘*koričnevij* suit’ is recorded but not ‘*buryj* suit’; similarly, Google returns *mužskoj koričnevij kostjum* ‘male *koričnevij* suit’ 1,680 times but returns no instances of *mužskoj buryj kostjum*. Thus, the predominant collocation of *koričnevij* with nouns for artefacts and of *buryj* with nouns for natural objects is in accord with the perception of *koričnevij* as a “pure”, homogeneous colour and of *buryj* as an “impure”, nonhomogeneous colour.

The linguistic behaviour of the two terms indicates that, in psycholinguistic terms, *buryj* appears to have emerged as a “proto-archaic” category encompassing both the colour solid “surface” of (relatively) saturated browns but also the “inner” desaturated “core” of the colour solid (cf. MacLaury 1992, 2007). Being applied linguistically to natural objects, *buryj* continues to retain its macrocolour sense denoting the BROWN and BROWN-GREY areas,

probably with a greater denotative volume and blurred category boundaries, as compared to the new ‘brown’ term.

The emergence of *buryj*’s “contestant”, *koričnevij*, was driven by technological developments in the dyeing industry, which enabled the production of artefacts with a “pure” brown appearance, and, correspondingly, by the communicative need in commerce and fashion. The incipient ‘brown’ term *koričnevij* was bestowed with a greater denotative precision – involving circumscribed denotata of saturated brown colours, thus, increasing the emphasis on the distinction from “impure” members of the BROWN category (cf. MacLaury’s [1991, 1992] scheme of BCC evolution in Supplement S1]. The “pure” hue tone implied by *koričnevij* was the principal reason for it gaining a more salient status in Russian colour vocabulary. *Koričnevij*, as well as collocating with nouns for more and more objects, also continues to increasingly encroach into the area of *buryj* denotative meanings, expanding from the colour space “skin” of saturated browns in the direction of the colour space “core” of less saturated browns.

Beyond Russian, we adduce parallels of similar lexical mutability of the BROWN area. Some other modern languages also manifest two lexical items for this category, with referential ranges greatly overlapping but with only one being a BCT. We are aware of three European languages in which an old ‘brown’ term is in the process of being replaced by a new one. In French, the historically older term *brun* is being supplanted by the more recent *marron* (Forbes 1979, 2006). In Spanish, the old term *pardo* ‘grey, brown, dusky, cloudy’ (Peers et al. 1959) was superseded by *marrón* ‘brown’, a BCT in modern Castilian Spanish. In Galician, *castaño* appears to be “falling out” of basicness, being used predominantly by the older generations, compared to younger generations who name brown *marrón* (a Spanish loanword), while reserving *castaño* for naming the colour of eyes and hair (Villanueva Gesteira 2009).

4.3 Connotative meanings of *koričnevij* and *buryj*

We would like to address the connotative aspect of collocations with “Russian browns” as revealed by an analysis of GBN bigrams, and which appears to be related to differences in the affective and aesthetic appraisal of the colour implied by each of the two terms. We observe that *koričnevij* is the term predominantly used with a neutral meaning for characterising

brown artefacts (e.g., *koričnevye tufli* ‘brown shoes’) or in positive contexts and, frequently, with diminutives (e.g., *malen’kaja koričnevaja sumočka* ‘a little brown bag’).

Buryj has a neutral meaning in naming the brown colour of natural objects (e.g., *buryj medved’* ‘brown bear’, *buryj ugol’* ‘brown coal’). However, in characterising the brown colour of artefacts it may have a pejorative nuance: it implies a colour that is no longer flawless, may appear repugnant, and manifests fading and the reduction in quality of an object (Hill 1972). GBN attests that *buryj* frequently collocates with nouns denoting foul, rotten objects, such as *gnil’* ‘putridity’ or *ržavčina* ‘rust’: in the GBN, the expression *buraja gnil’* is recorded 8 times more frequently than *koričnevaja gnil’*.

On the Internet, the expression *prijatnyj koričnevyy* ‘pleasant *koričnevyy*’ was recorded 10 times more frequently than *prijatnyj buryj*; conversely, the record of *neprijatnyj buryj* ‘unpleasant *buryj*’ was five times higher than that of *prijatnyj buryj*. Furthermore, in our recent (unpublished) study¹, using the neuronal network BERT, of the affective meanings of contexts in which Russian colour terms were used, we found that on a 10-point “pleasantness” scale, *koričnevyy* was rated 6.4, and *buryj* 4.8, the lowest among frequent CTs. Notably, GBN examples also provide evidence that when a negative context is involved, the nouns denoting artefacts collocate with *buryj* (e.g., factory chimneys exuding soot or black smoke; worn clothes; or drab old human skin; Supplements SE21–SE24).

Markedly, all *buryj*-derived inceptive verbs – *buret’* ‘to acquire a *buryj* shade’, *zaburet’* ‘to begin acquiring a *buryj* shade’, *poburet’* ‘becoming *buryj*’ – may indicate a natural process of grain ripening (e.g., rye crop maturation in the fields) (cf. Bakhilina 1975: 223). More frequently, however, they involve a process of the object’s degradation caused by its age or lengthy exposure to sunlight or rain (e.g., the decay of the logs of a wooden house; see also examples in Supplement SE2). In relation to artefacts, these verbs, collectively, have negative connotations and convey the regression of an originally “hue-circumscribed” colour towards an “impure”, “dirty”, “hue-undefined” colour (Bakhilina 1975: 224).

1. <https://kpfu.ru/tehnologiya-sozdaniya-semanticheskikh-elektronnyh.html>

Interestingly, according to the NRC, from the mid-20th century onwards, the verb *zaburet'* has been used predominantly in a metaphoric sense about a person who becomes priggish, although Bakhilina (1975: 223) recorded this meaning as dialectal only (in the Tula region), where it means “to give oneself airs, to redden while aware of one’s importance.

Koričnevij-derived verbal forms, analogous to those of *buryj*, emerged only recently (as Internet-vocabulary innovations): *koričnevet'*, *pokoričnevet'* and *zakoričnevet'* ‘to acquire a brown tint’ refer to a brownish suntan or a brown crust acquired during the cooking or baking process, all evoking positive connotations (see examples in Supplement SE3). It is noteworthy that these morphological variants are similar to verbal forms of other established Russian BCTs, such as *golubet'* ‘having a light blue appearance’ (e.g., in relation to the sky), *zazelenet'* ‘becoming green’ (about trees in spring), or *pobelet'* ‘becoming white’.

Thus, the above overview of contexts provides evidence that, along with the predominant differentiation of the pattern of *koričnevij*- and *buryj*-collocations summarised in §4.1, the contexts demarcate two further ways to discern the “Russian browns” – denotative, “pure” vs. “impure”, and connotative, “pleasant” vs. “unpleasant”, respectively.

4.4 The *koričnevij* term as a socio-cultural marker

We conclude the Discussion by quoting Biggam (2006: 169–170): “The questions how? and when? have been addressed, but there remains the most curious question of all: why?”. Her case study pertains to an analysis of historical Old English texts: the replacement of *hæwen*, the term that signified a macro-category of BLUE, GREY and GREY-BLUE, by *bleu*, adopted in the 11th and 12th centuries after the Norman Conquest. When it was first introduced, *bleu* did not have a prominent blue sense but indubitably developed it later and became the superordinate for the ‘blues’. Biggam (2006: 170) argues that “... the reason may be found in the semantic field of cloth and clothing”.

We are inclined to share Biggam’s viewpoint: indeed, in the earliest records of *koričnevij* in its colour meaning in the 17th-century trade sources (Bakhilina 1975: 228–229; Russian National Corpus n.d.) the term collocates with nouns denoting dyed clothes (*barxat* ‘velvet’, *sukno* ‘cloth’, *atlas* ‘satin’, *tafta* ‘taffeta’, *tkan'* ‘fabric’), clothing (*kaftan* ‘gabardine suit, surcoat’, *šal'* ‘shawl’), and furs (e.g., *sobolinaja šapka* ‘sable hat’). It is plausible that the new term emerged with the demand for the denotative discernment of the two ‘brown’

terms, stipulated by changes in the Russian material culture of the 17th century, to refer to vivid and homogeneous brown shades of the fabrics dyed in a more elaborate and expensive way, compared to a dull brown implied by *buryj*. (For historical details of introducing cinnamon in Russia and *koričnevij* entrenching, see Supplement S5.)

Notably, the clothing items listed above hint that in its semantic development, *koričnevij* acquired the role of a social identity marker: it collocates with nouns signifying furs and high-quality luxurious garments of the Russian nobility. It is conceivable that the nobility introduced and cemented the “branding” term of their expensive cloths by bestowing prestige on the more recently named colour and, thus, conveying their affluence and political power to the reader/interlocutor (cf. MacLaury 1991, 1992; Swearingen 2014).

5. Conclusions

We investigated the linguistic mechanisms of colour term evolution from a diachronic perspective. For this, we explored the linguistic behaviour of the two “Russian browns”, “old” (declining in use) *buryj*, and “new” *koričnevij*, a basic term in modern Russian. For computational analysis, the third version of the GBN Russian subcorpus was used spanning Russian books published between 1486 and 2019. We ascertained the frequency of occurrence of each term and the dynamics of the terms’ collocations with nouns signifying specified classes of objects.

The analysis allowed us to confirm Rakhilina’s (2007a, 2007b, 2008) hypothesis that, in the process of the coexistence of two terms in a language which linguistically compete for a certain area of colour space, an old term’s use dwindles to apply predominantly to natural objects, whereas an incipient colour term initially collocates with nouns denoting artefacts but gradually expands to the realm of natural objects (see also Rakhilina & Paramei 2011).

Colour characteristics of the objects denoted by the two ‘brown’ terms indicate that, from its incipience, the new term is not a full synonym of the old macro-category (cf. Biggam 2012): whereas *buryj* appears to denote a broad variety of brown colours, including greyish shades (the “core” of a colour solid), the denotative meaning of *koričnevij* is more circumscribed, and characterises homogeneous and saturated colours of the BROWN area. This observation is in line with MacLaury’s (1991, 1992) scheme of BCC evolution, according to which the emerging colour term implies emphasis on its distinction from “impure” members of the “proto-archaic” (here: *buryj*) category.

The novelty of the present study is threefold:

1. Our analysis showed that linguistic entrenchment of the incipient colour term *koričnevijj*, which supplants *buryj* as basic, varies depending on the nominal field: natural objects, artefacts, abstract concepts and figurative expressions. Specifically, although the rate of transition from *buryj* to *koričnevijj* is comparable for distinct noun categories, the difference between them is in the time-point of such a transition: front runners in collocating with *koričnevijj* are nouns for artefacts; this is followed by abstract concepts and polysemous nouns denoting both natural objects and artefacts, while nouns for natural objects lag behind all other categories.
2. Linguistically in Russian, natural objects “become” *koričnevijj* in the contexts where they acquire the meaning of artefacts – animals represented in depictions and applications; as toys, national symbols and fictional characters (e.g., a brown bear); or minerals in their processed state (e.g., brown diamond). Furthermore, in modern Russian, *koričnevijj* is predominantly used in brown translations of zoological terms and abstract concepts (e.g., brown scum; brown zone).
3. Based on the analysis of frequencies of occurrence and collocational patterns, we estimated that in Russian, *koričnevijj* supplanted *buryj* as a basic term naming the BROWN colour space area around 100 years ago. We conjecture that this process was instigated by globalised technological developments in the dyeing industry and, specifically for the Russian language, by a post-October Revolution (1917) substantial increase in the production and dissemination of (mainly new) books.

The linguistic refinement of the BROWN area in Russian manifests some general mechanisms of an increasingly fine-grained lexical partition of the colour space: it reflects a confluence between psychology, language and cultural-economic factors that together encourage differentiation in colour terminology by raising the level of chromatic precision under the need for communication efficiency (Steels & Belpaeme 2005; Conway et al. 2020; Zaslavsky et al. 2020, 2022).

Moreover, diachronic distributional analysis of *buryj* and *koričnevijj*, and their distinct connotative meanings in modern Russian corroborate Kul’pina’s (2019) conjecture that, in addition to signifying colour as such, colour terms are influenced by extralinguistic factors, and acquire and convey ethno-specific aesthetic and axiological meanings.

Funding

Bochkarev, Shevlyakova, Solovyev: Russian Science Foundation project “Distributional-quantitative analysis of semantic changes based on large diachronic text corpora”, Grant No. 20-18-00206.

Acknowledgements

The authors are grateful to Susanne Vejdemo for valuable comments on the semasiological evolution of colour terms. We greatly appreciate the valuable comments of Carole Biggam and another, anonymous, reviewer, which were very helpful for improving the original version of the paper. Our sincere thanks go to two native speakers, Ingrid Boedker and the expert in colour naming, for thorough proofreading of a revised version of the manuscript. We also are thankful for the help of Valérie Bonnardel and Martin Blasczyk for translation of the Abstract into French and German respectively.

Abbreviations

BCC	basic colour category
BCT	basic colour term
<i>b/k</i> -nouns	nouns that combine with either <i>buryj</i> or <i>koričnyj</i>
CT	colour term
CTIT	characteristic time interval of transition from <i>buryj</i> to <i>koričnyj</i> in collocations with nouns
GBN	Google Books Ngram

References

- Bakhilina, Natalia B. 1975. *Istorija cvetooboznačenij v ruskom jazyke* [History of colour terms in Russian]. Moscow: Nauka (in Russian).
- Berlin, Brent and Paul Kay. 1969/1991. *Basic color terms: Their universality and evolution*. Berkeley, CA: University of California Press.
- Biggam, Carole P. 2006. Political upheaval and disturbance in the colour vocabulary of Early English. In Carole P. Biggam and Christian Kay (eds.), *Progress in colour studies. Vol. I. Language and culture*, 159–179. Amsterdam/Philadelphia: John Benjamins.
- Biggam, Carole P. 2012. *The semantics of colour: A historical approach*. Cambridge, UK: Cambridge University Press.
- Bocharov, Victor V., Svetlana V. Alexeeva, Dmitry V. Granovsky, Ekaterina V. Protopopova et al. 2013. Crowdsourcing morphological annotation. *Computational Linguistics and Intellectual Technologies* 13. 109–114. http://www.dialog-21.ru/media/1308/dialog_2013_vol1web.pdf
- Bochkarev, Vladimir V., Anna V. Shevlyakova, Galina V. Paramei, and Ekaterina V. Rakhilina. 2021. A quantitative study of Russian colour terms *buryj* and *koričnevij* in the Google Books Ngram corpus. *Linguistic Forum 2020: Language and Artificial Intelligence, Moscow 12-14 November 2020. CEUR Workshop Proceedings* 2852. <http://ceur-ws.org/Vol-2852/>
- Brown, Peter F., Stephen A. Della Pietra, Vincent J. Della Pietra, Jennifer C. Lai, and Robert L. Mercer. 1992. An estimate of an upper bound for the entropy of English. *Computational Linguistics* 18(1). 31–40. <https://dl.acm.org/doi/abs/10.5555/146680.146685>
- Chaabouni, Rahma, Eugene Kharitonov, Emmanuel Dupoux, and Marco Baroni. 2021. Communicating artificial neural networks develop efficient color-naming systems. *Proceedings of the National Academy of Sciences of the U.S.A.* 118(12). e2016569118. <https://doi.org/10.1073/pnas.2016569118>
- Conway, Bevil R., Sivalogeswaran Ratnasingam, Julian Jara-Ettinger, Richard Futrell, and Edward Gibson. 2020. Communication efficiency of color naming across languages provides a new framework for the evolution of color terms. *Cognition* 195. 104086. <https://doi.org/10.1016/j.cognition.2019.104086>
- Corbett, Greville G. and Ian R.L. Davies. 1995. Linguistics and behavioural measures for ranking basic colour terms. *Studies in Language* 19(2). 301–357. <https://doi.org/10.1075/sl.19.2.02cor>
- Corbett, Greville G. and Gerry Morgan 1988. Colour terms in Russian: Reflections of typological constraints in a single language. *Journal of Linguistics* 24(1). 31–64. <https://www.jstor.org/stable/4175920>
- Dal, Vladimir. 1863–1866. *Tolkovyj slovat' živogo velikoruskogo jazyka* [Explanatory dictionary of the living Great Russian language]. St. Petersburg: Imperial Academy of Sciences (in Russian).
- Davies, Ian R. L. and Greville G. Corbett. 1994. The basic color terms of Russian. *Linguistics* 32(1). 65–89. <https://doi.org/10.1515/ling.1994.32.1.65>
- Decock, Lieven. 2021. Conceptual change and conceptual engineering: the case of colour concepts. *Inquiry* 64(1-2). 168–185. <https://doi.org/10.1080/0020174X.2020.1784783>
- Endres, Dominik M. and Johannes E. Schindelin. 2003. A new metric for probability distributions. *IEEE Transactions on Information Theory* 49(7). 1858–1860. doi: 10.1109/TIT.2003.813506
- Falla, Paul S. 1984. *The Oxford English–Russian dictionary*. Oxford: Clarendon Press.

- Fano, Robert M. 1961. *Transmission of information: A statistical theory of communication*. Cambridge, MA: MIT Press.
- Forbes, Isabel. 1979. The terms *brun* and *marron* in modern standard French. *Journal of Linguistics* 15(2). 295–305. <https://doi.org/10.1017/S0022226700016406>
- Forbes, Isabel. 2006. Age-related differences in the basic colour vocabulary of French. In Carole P. Biggam and Christian J. Kay (eds.), *Progress in colour studies. Vol. 1: Language and culture*, 101–109. Amsterdam/Philadelphia: John Benjamins.
- Frumkina, Rebecca M. 1984. *Cvet, smysl, srodstvo. Aspekty psixolingvističeskogo analiza* [Colour, meaning, and similarity: Aspects of a psycholinguistic analysis]. Moscow: Nauka (in Russian).
- Frumkina, Rebecca M. and Alexei V. Mikhejev, 1996. *Meaning and categorization*. New York: Nova Science.
- Gage, John. 1999. What meaning had colour in early societies? *Cambridge Archaeological Journal* 9. 109–126. <https://doi.org/10.1017/S0959774300015237>
- Gibbons, Jean D. and Subhabrata Chakraborti. 2010. *Nonparametric statistical inference*. 5th ed. Boca Raton, FL: Taylor & Francis/CRC Press.
- Gippius, Alexei A. 2020. Berestjanye gramoty iz raskopok 2019 g. v Velikom Novgorode i Staroj Russe [Birchbark letters from Veliky Novgorod and Staraya Russa excavations of 2019]. *Voprosy Jazykoznanija*, Iss. 5, 22–37 (in Russian).
- Google Books Ngram Viewer. 2020. <https://books.google.com/ngrams>
- Greenfeld, Philip J. 1986. What is grey, brown, pink, and sometimes purple: The range of "wild-card" color terms. *American Anthropologist* 88(4). 908–916. <https://www.jstor.org/stable/679079>
- Gulordava, Kristina and Marco Baroni. 2011. A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus. In *Proceedings of the GEMS 2011 Workshop on Geometrical Models of Natural Language Semantics*. 67–71. <https://aclanthology.org/W11-2508>
- Haynie, Hannah J. and Claire Bower. 2016. Phylogenetic approach to the evolution of color term systems. *Proceedings of the National Academy of Science of the U.S.A.* 113(48). 13666–13671. <https://doi.org/10.1073/pnas.1613666113>
- Heaps, Harold S. 1978. *Information retrieval: Computational and theoretical aspects*, 206–208. New York: Academic Press.
- Herne, Gunnar. 1954. *Die Slavischen Farbenbenennungen. Eine semasiologische-etymologische Untersuchung*. Uppsala: Almqvist & Wiksells.
- Hill, Peter M. 1972. *Die Farbwörter der russischen und bulgarischen Schriftsprachen der Gegenwart (Versuch einer Klassifikation und einer strukturell-semantischen Analyse)*. Amsterdam: Adolph M. Hakkert.
- Hilpert, Martin and Stefan T. Gries. 2009. Assessing frequency changes in multistage diachronic corpora: Application for historical corpus linguistics and the study of language acquisition. *Literary and Linguistic Computing* 24(4). 385–401. <https://doi.org/10.1093/lc/fqn012>
- Hilpert, Martin and Stefan T. Gries. 2016. Quantitative approaches to diachronic corpus linguistics. In Merja Kytö and Päivi Pahta (eds.), *The Cambridge handbook of English historical linguistics*, 36–53. Cambridge, UK: Cambridge University Press.
- Holland, Paul W. and Roy E. Welsch. 1977. Robust regression using iteratively reweighted least-squares. *Communications in Statistics – Theory and Methods* 6(9). 813–827. <https://doi.org/10.1080/03610927708827533>

- Hollander, Myles and Douglas A. Wolfe. 1999. *Nonparametric statistical methods*. Hoboken, NJ: John Wiley & Sons.
- Jameson, Kimberly A. and Roy G. D'Andrade. 1997. It's not really red, green, yellow, blue: An inquiry into cognitive color space. In Clyde L. Hardin & Luisa Maffi (eds.), *Color categories in thought and language*, 295–318. Cambridge, MA: Cambridge University Press.
- Joula, Patrick. 2003. The time course of language change. *Computers and the Humanities* 37. 77–96. <https://doi.org/10.1023/A:1021839220474>
- Jraissati, Yasmina and Igor Douven. 2017. Does optimal partitioning of color space account for universal color categorization? *PLoS ONE* 12(6). e0178083. <https://doi.org/10.1371/journal.pone.0178083>
- Kay, Paul. 2015. Universality of color categorization. In Andrew J. Elliot, Mark D. Fairchild, and Anna Franklin (eds.), *Handbook of color psychology*, 245–258. Cambridge, UK: Cambridge University Press.
- Kay, Paul and Luisa Maffi. 1999. Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist* 101. 743–760. <https://doi.org/10.1525/aa.1999.101.4.743>
- Koplenig, Alexander. 2017. The impact of lacking metadata for the measurement of cultural and linguistic change using the Google Ngram data sets—Reconstructing the composition of the German corpus in times of WWII. *Digital Scholarship in the Humanities* 32(1). 169–188. <https://doi.org/10.1093/llc/fqv037>
- Koptjevskaja-Tamm, Maria, Ekaterina Rakhilina, and Martine Vanhove. 2015. The semantics of lexical typology. In Nick Riemer (ed.), *The Routledge handbook of semantics*, 434–454. London & New York: Routledge.
- Krivko, Roman. 2021. Old East Slavic. In Carole P. Biggam and Kirsten Wolf (eds.), *A cultural history of color in the Medieval Age*, 124–128. London: Bloomsbury.
- Kulkarni, V., Al-Rfou, R., Perozzi, B., & Skiena, S. (2015). Statistically significant detection of linguistic change. In *Proceedings of the 24th International Conference on World Wide Web, May 2015, Florence* (pp. 625–635). <https://doi.org/10.1145/2736277.2741627>
- Kul'pina, Valentina G. 2001. *Lingvistika cveta. Terminy cveta v pol'skom i ruskom jazykax* [Linguistics of colour. Colour terms in Polish and Russian]. Moscow: Russkie slovari (in Russian).
- Kul'pina, Valentina G. 2007. Sistema cvetooboznačenij ruskogo jazyka [The system of Russian colour terms]. In Aleksandr P. Vasilevich (ed.), *Naimenovanija cveta v indoevropskix jazykax: Sistemnyj i istoričeskij analiz* [Colour terms in Indo-European languages: Systematic and historical analysis], 126–184. Moscow: KomKniga (in Russian).
- Kul'pina, Valentina G. 2019. *Lingvističeskaja cvetologija: ot istorii k sovremennosti cvetovyx konceptosfer* [Linguistic colour theory: From history to modernity of coloristic conceptual spheres]. Moscow: MAKS Press (in Russian).
- Levinson, Stephen C. 2000. Yélî Dnye and the theory of basic color terms. *Journal of Linguistic Anthropology* 10(1). 3–55. <https://doi.org/10.1525/jlin.2000.10.1.3>
- Lin, Yuri, Jean-Baptiste Michel, Erez L. Aiden, Jon Orwant, Will Brockman, and Slav Petrov. 2012. Syntactic annotations for the Google Books Ngram corpus. In *Proceedings of the ACL 2012 System Demonstrations*, 169–174. Association for Computational Linguistics. <https://aclanthology.org/P12-3029>
- Lopatin, Vladimir V. 2013. *Tolkovyj slovar' sovremennogo ruskogo jazyka* [A glossary of modern Russian]. Moscow: Èksmo (in Russian).

- Lucy, John A. 1997. The linguistics of “color”. In Clyde L. Hardin and Luisa Maffi (eds.), *Color categories in thought and language*, 320–346. Cambridge, UK: Cambridge University Press.
- Lyashevskaya, Olga N. and Sergei A. Sharov. 2009. *Častotnyj slovar’ sovremennogo russkogo jazyka (na naterialax Nacional’nogo korpusa russkogo jazyka)* [Frequency dictionary of the modern Russian language (the Russian National Corpus)]. Moscow: Azbukovnik (in Russian).
- Lyons, John. 1995. Colour in language. In Trevor Lamb, and Janine Bourriau (eds.), *Colour: Art and science*, 194–234. Cambridge, UK: Cambridge University Press.
- MacLaury, Robert E. 1991. Social and cognitive motivations of change: Measuring variability in color semantics. *Language* 67(1). 34–62. <https://www.jstor.org/stable/415538>
- MacLaury, Robert E. 1992. From brightness to hue: An explanatory model of color-term evolution [and comments and reply]. *Current Anthropology* 33(2). 137–186. <http://www.jstor.com/stable/2743839>
- MacLaury, Robert E. 1997. *Color and cognition in Mesoamerica: Constructing categories as vantages*. Austin: University of Texas Press.
- MacLaury, Robert E. 2007. Categories of desaturated-complex color. Sensorial, perceptual, and cognitive models. In Robert E. MacLaury, Galina V. Paramei, and Don Dedrick (eds.), *Anthropology of color: Interdisciplinary multilevel modeling*, 125–150. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/z.137.10mac>
- Michel, Jean-Baptiste, Yuan K. Shen, Aviva P. Aiden, Adrian Veres, A. et al. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331(6014). 176–182. <https://doi.org/10.1126/science.1199644>
- Morgan, Gerry and Greville Corbett. 1989. Russian colour term salience. *Russian Linguistics* 13(2). 125–141. <https://www.jstor.org/stable/40160229>
- Nicklas, Charlotte. 2017. New words and fanciful names: Dyes, color and fashion in the mid-nineteenth century. In Regina L. Blaszczyk and Uwe Spiekermann (eds.), *Bright modernity. Color, commerce and consumer culture*, 97–111. New York: Palgrave MacMillan.
- OpenCorpora, n.d. <http://opencorpora.org/dict.php>
- Paramei, Galina V. and David L. Bimler. 2021. Language and psychology. In Anders Steinvall and Sarah Street (eds.), *A cultural history of color. The Modern Age: From 1920 to present*, 117–134. London: Bloomsbury.
- Paramei, Galina V., Yulia A. Griber, and Dimitris Mylonas. 2018. An online color naming experiment in Russian using Munsell color samples. *Color Research & Application* 43(3). 358–374. <https://doi.org/10.1002/col.22190>
- Pechenick, Eitan A., Christopher M. Danforth, and Peter S. Dodds. 2015. Characterizing the Google Books corpus: Strong limits to inferences of socio-cultural and linguistic evolution. *PLoS ONE* 10(10). e0137041. <https://doi.org/10.1371/journal.pone.0137041>
- Peers, E. Allison, José V. Barragán, Francesco A. Vinyals, and Jorge A. Mora. (eds.). 1959. *Cassell's Spanish Dictionary: Spanish-English, English-Spanish*. New York: Funk & Wagnalls.
- Rakhilina, Ekaterina V. 2007a. Linguistic construal of colors: The case of Russian. In Robert E. MacLaury, Galina V. Paramei, and Don Dedrick (eds.), *Anthropology of color: Interdisciplinary multilevel modeling*, 363–377. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/z.137.24rak>
- Rakhilina, Ekaterina V. 2007b. O semantike prilagatel’nyx cveta [On semantics of colour

- adjectives]. In Aleksandr P. Vasilevich (ed.), *Naimenovanija cveta v indoevropskix jazykax: Sistemnyj i istoričeskij analiz* [Colour terms in Indo-European languages: Systematic and historical analysis], 29–39. Moscow: KomKniga (in Russian).
- Rakhilina, Ekaterina V. 2008. *Kognitivnyj analiz predmetnyx imen: semantika i sočetaemost'* [Cognitive analysis of object names: Semantics and combinability]. 2nd ed. Moscow: Russkie slovari (in Russian).
- Rakhilina, Ekaterina V. and Galina V. Paramei. 2011. Colour terms: Evolution via expansion of taxonomic constraints. In Carole P. Biggam, Carole A. Hough, Christian J. Kay, and David R. Simmons (eds.), *New directions in colour studies*, 121–131. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/z.167.15rak>
- Regier, Terry, Paul Kay, and Khetarpal. 2007. Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences of the U.S.A.* 104(4). 1436–1441. <https://doi.org/10.1073/pnas.0610341104>
- Richey, Sean and J. Benjamin Taylor. (2020). Google Books Ngrams and political science: Two validity tests for a novel data source. *PS: Political Science & Politics* 53(1). 72–77. doi:10.1017/S1049096519001318
- Russian National Corpus, n.d. <https://ruscorpora.ru/en/>
- Safuanova, Olga V. and Nina N. Korzh. 2007. Russian color names. Mapping into a perceptual color space. In Robert E. MacLaury, Galina V. Paramei, and Don Dedrick (eds.), *Anthropology of color: Interdisciplinary multilevel modeling*, 55–74. Amsterdam/Philadelphia: John Benjamins.
- Saunders, Barbara A. C. and Jan van Brakel. 1997. Are there non-trivial constraints on colour categorisation? *Behavioral and Brain Sciences* 20(2). 167–228. <https://doi.org/10.1017/S0140525X97531426>
- Slovari russkogo jazyka 11–17 vekov: Ètimologija i istorija slov russkogo jazyka* [Dictionaries of the Russian language of the 11th–17th centuries: Etymology and history of Russian words]. 1975–. Moscow: Nauka (in Russian).
- Slovar' Akademii Rossijskoj* [Dictionary of the Russian Academy] (1789–1794). St. Petersburg: Imperial Academy of Sciences (in Russian).
- Solovyev, Valery D., Vladimir V. Bochkarev, and Svetlana S. Akhtyamova. 2020. Google Books Ngram: Problems of representativeness and data reliability. In Alexander Elizarov, Boris Novikov, and Sergey Stupnikov (eds.), *Data Analytics and Management in Data Intensive Domains. DAMDID/RCDL 2019. Communications in Computer and Information Science* Vol. 1223, 147–162. Cham: Springer. https://doi.org/10.1007/978-3-030-51913-1_10
- Steels, Luc. 2011. Modeling the cultural evolution of language. *Physics of Life Reviews* 8(4). 339–356. <https://doi.org/10.1016/j.plrev.2011.10.014>
- Steels, Luc and Tom Belpaeme. 2005. Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences* 28. 469–489. <https://doi.org/10.1017/S0140525X05000087>
- Steinvall, Anders. 2002. *English colour terms in context*. PhD thesis. Umeå University.
- Street, James O., Raymond J. Carroll, and David Ruppert. 1988. A note on computing robust regression estimates via iteratively reweighted least squares. *The American Statistician* 42(2). 152–154. <https://doi.org/10.2307/2684491>
- Swearingen, Andrew. 2014. From blood to worms: The semantic evolution of a Portuguese colour term. In Wendy Anderson, Carole P. Biggam, Carole Hough, and Christian Kay (eds.), *Colour studies: A broad spectrum*, 79–92. Amsterdam/Philadelphia: John

Benjamins.

- Tahmasebi, Nina, Lars Borin, and Adam Jatowt 2021. Survey of computational approaches to lexical semantic change detection. In Nina Tahmasebi, Lars Borin, Adam Jatowt, Yang Xu, and Simon Hengchen (eds.), *Computational approaches to semantic change*, 1–9. Berlin: Language Science Press. <https://langsci-press.org/catalog/book/303>
- Tang, Xuri. 2018. A state-of-the-art of semantic change computation. *Natural Language Engineering* 24(5). 649–676. doi: 10.1017/S1351324918000220
- Ulukhanov, Igor S. 2015. *Glagol'noe slovoobrazovanie sovremennogo russkogo jazyka* [Adverbial derivation of modern Russian]. Vol. I. *Glagoly, motivirovannye imenami i meždometijami* [The verbs motivated by proper names and interjections]. Moscow: Azbukovnik (in Russian).
- van Leijenhorst, Dick C., and Theo P. van der Weide. 2005. A formal derivation of Heaps' Law. *Information Sciences* 170(2-4). 263–272. <https://doi.org/10.1016/j.ins.2004.03.006>
- Vasilevich, Alexandr P., Svetlana N. Kuznecova, and Sergei S. Mishchenko. 2016. *Cvet i nazvanija cveta v rusском jazyke* [Colour and colour terms in the Russian language]. Moscow: LCI Publisher (in Russian).
- Villanueva Gesteira, María D. 2009. O campo semántico das cores visto por tres xeracións [The semantic domain of colours viewed by three generations]. *Estudos de Lingüística Galega* 1. 169–188 (in Galician).
- Werner, Heinz. 2004. *Comparative psychology of mental development* (with a new prologue by Margery B. Franklin). Clinton Corners, NY: Percheron Press.
- Wierzbicka, Anna. 2005. There are no "color universals" but there are universals of visual semantics. *Anthropological Linguistics* 47. 217–244. <https://www.jstor.org/stable/25132327>
- Worth, Dean S., Andrew S. Kozak, and Donald B. Johnson. 1970. *Russian derivational dictionary*. New York: Elsevier.
- Zaslavsky, Noga, Charles Kemp, Naftali Tishby, and Terry Regier. 2020. Communicative need in color naming. *Cognitive Neuropsychology* 37. 312–324. <https://doi.org/10.1080/02643294.2019.1604502>
- Zaslavsky, Noga, Karee Garvin, Charles Kemp, Naftali Tishby, and Terry Regier. 2022. The evolution of color naming reflects pressure for efficiency: Evidence from the recent past. *Journal of Language Evolution*. Early view. Izac001. <https://doi.org/10.1093/jole/lzac001>

Résumé

A l'aide du sous-corpus russe (2020) de Google Books Ngram, nous avons étudié la diachronie de la sémantique distributionnelle de deux termes de couleur (TC) concurrents, significatif 'marron', en langue russe: *buryj* (XI^e siècle) et *koričnevij* (XVII^e siècle). L'analyse de bigrammes en séries chronologiques (1800–2019) mesure les fréquences d'occurrence de chaque terme ainsi que les changements de combinabilité avec des noms d'objets naturels, d'artefacts, de concepts abstraits et d'expressions figuratives. Dans les années 1920, la fréquence de *koričnevij* était supérieure à celle de *buryj*, confirmant son statut de TC de base en russe moderne. L'indice de perplexité indique que *koričnevij* a régulièrement augmenté le nombre d'objets ainsi désignés, les artefacts étant les premiers dans la transition de *buryj* à *koričnevij*. Les résultats confirment l'hypothèse de Rakhilina (2007a, 2007b, 2008) selon laquelle un TC émerge initialement par sa colocalisation avec des noms d'artefacts, mais s'étend ensuite progressivement au domaine des objets naturels supplantant ainsi un ancien TC. De plus, *koričnevij* et *buryj* se distinguent par des dénnotations et des connotations. Nos résultats donnent un aperçu des mécanismes généraux de l'évolution linguistique sur l'émergence d'un TC de base.

Zusammenfassung

Wir untersuchten die Diachronie der distributionalen Semantik der zwei konkurrierenden russischen Farbbezeichnungen (FB) für 'braun', *buryj* (11. Jahrhundert) und *koričnevij* (17. Jahrhundert) mittels Google Books Ngram (2020). Die Zeitreihenanalyse (1800–2019) der Bigramme erlaubte, die Auftretenshäufigkeit von jedem dieser FB zu messen, wie auch die Änderungen der Kombinierbarkeit dieser mit Substantiven für natürliche Objekte, Artefakte (menschengemachte Gegenstände), abstrakte Konzepte und figurative Ausdrücke. Bezüglich der Häufigkeit wurde *buryj* von *koričnevij* in den 1920er Jahren überholt, was den Grundstatus von *koričnevij* im modernen Russisch bestätigt. Der Perplexitätsindex weist darauf hin, dass *koričnevij* den Bereich der bezeichneten Objekte stetig vergrößerte, wobei Artefakte die Spitzenreiter beim Übergang von *buryj* zu *koričnevij* sind. Die Ergebnisse bestätigen die Hypothese von Rakhilina (2007a, 2007b, 2008), dass eine neu erscheinende FB anfangs mit Substantiven kolloziert, die Artefakte bezeichnen, allmählich aber in den Bereich der natürlichen Objekte expandiert, die alte FB verdrängend. Darüber hinaus

unterscheiden sich *koričevyj* und *buryj* durch die denotativen und konnotativen Bedeutungen. Die Ergebnisse liefern Einblicke in generelle Mechanismen der linguistischen Evolution einer emergenten FB, die den Grundstatus gewinnt.

Address for correspondence

Galina V. Paramei
Department of Psychology
Liverpool Hope University
Hope Park
L16 9JD Liverpool
United Kingdom

parameg@hope.ac.uk
<http://orcid.org/0000-0003-2611-253X>

Co-author information

Vladimir V. Bochkarev
Institute of Physics
Kazan Federal University
Kremlyovskaya Str. 18
420008 Kazan
Russia

vbochkarev@mail.ru
<https://orcid.org/0000-0001-8792-1491>

Anna V. Shevlyakova
Institute of Philology and Intercultural Communication
Kazan Federal University
Tatarstan Str. 2
420021 Kazan
Russia

anna_ling@mail.ru
<https://orcid.org/0000-0002-2659-1887>

Valery D. Solovyev
Institute of Philology and Intercultural Communication
Kazan Federal University
Kremlyovskaya Str. 18
420008 Kazan
Russia

maki.solovyev@mail.ru
<https://orcid.org/0000-0003-4692-2564>

Ekaterina V. Rakhilina
School of Linguistics
Higher School of Economics
Staraya Basmannaya Str. 21/4, Bld. 1
105066 Moscow
Russia

rakhilina@gmail.com

<https://orcid.org/0000-0002-7126-0905>

Supplementary Materials

Supplement S1: MacLaury's 4-stage scheme of the evolutionary dynamics of BCCs

Supplement S2: MacLaury's trajectory of BCC evolution in computer simulations

Supplement S3. History of the colour term *buryj* (Bakhilina 1975: 227–228)

Supplement S4: Semantic cluster of most frequent terms for 'brown' in modern Russian

Supplement S5: Historical details of the introduction of cinnamon in Russia and *koričnevij* entrenching

Supplement Table S1

Supplement Figures S1–S3

Supplementary Examples (SE) of *buryj*- and *koričnevij*-collocations

Supplement S1: MacLaury's 4-stage scheme of the evolutionary dynamics of BCCs

MacLaury (1997, 2007) offered a 4-stage scheme of decomposing the brightness-based “proto-category” (illustrated in pictorial terms by Vejdemo 2018: Figure 1). At stage 1, in relation to the old term, an incipient term originally manifests synonymy. At stage 2, this slowly transforms into co-extension: the two terms retain denotative (near-)synonymy, but their ‘best examples’ diverge, the process stipulated by the speaker’s vantage (dominant or recessive), i.e. pragmatic and semantic context, whereby the two vantages are named separately. At stage 3, an inclusion relationship follows, with the incipient term carving a certain denotative sub-range around its ‘best example’ abutting the shrunken range of the old term. Finally, at stage 4, complementation signifies full denotative divorce of the two terms, whereby the now denotatively circumscribed and linguistically unambiguously labelled new term expands and entrenches at the expense of the old term, whose denotative range shrinks.

Of note, MacLaury (1991: 57--58) singled out the rare case of the Tzotzil language, an exception that “stands sharply apart from the usual process” and reveals just a fragment of the full four-stage model that comprises solely its two last stages. In particular, in relation to the archaic term, the “new” colour term starts not as a synonym but as a restricted hyponym of the superordinate term (stage 3); due to social convention, the archaic term is maintained at a very low salience, whereas the incipient term becomes highly salient and fully-fledged in its own right (stage 4).

In an attempt to tackle the problem of the linguistic establishment of a novel CT, Vejdemo (2018) offers a semasiological approach to analysing corpora by further elaborating MacLaury’s (1997) 4-stage scheme. Of particular relevance in the present context – of the old Russian term for the ‘brown’ CC having been supplanted by the new one – is her observation of two instances of lexical replacement in Swedish names of BCCs recorded intergenerationally: in the PINK colour space area, the older generations’ *skär* is being supplanted by the younger generations’ *rosa*; similarly, in the PURPLE colour space area, *violett* and *gredelin* are being replaced by *lila*. In both cases, lexical replacement is accompanied by a denotative shift of semantic material to lighter shades.

References

MacLaury, Robert E. 1991. Social and cognitive motivations of change: Measuring variability

- in color semantics. *Language* 67(1). 34–62. <https://www.jstor.org/stable/415538>
- MacLaury, Robert E. 1997. *Color and cognition in Mesoamerica: Constructing categories as advantages*. Austin: University of Texas Press.
- MacLaury, Robert E. 2007. Categories of desaturated-complex color. Sensorial, perceptual, and cognitive models. In Robert E. MacLaury, Galina V. Paramei, and Don Dedrick (eds.), *Anthropology of color: Interdisciplinary multilevel modeling*, 125–150. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/z.137.10mac>
- Vejdemo, Susanne. 2018. Lexical change often begins and ends in semantic peripheries. Evidence from color linguistics. *Pragmatic & Cognition* 25(1). 50–85. <https://doi.org/10.1075/pc.00005.vej>

Supplement S2: MacLaury's trajectory of BCC evolution in computer simulations

Using evolutionary language games with artificial agents in computer simulations, Steels (2011) demonstrated the plausibility of the MacLaury alternative trajectory, as well as novelty and innovation as the driving force for achieving communication efficiency. In the same vein, in their discrimination–similarity game simulating the evolutionary dynamics of colour categorisation in the WCS, Komarova et al. (2007: 359) conclude that BCC evolution “is a combination of a minimal perceptual psychology of discrimination, simple pragmatic constraints involving communication, and simple learning rules”. The authors demonstrated the essential role of exogenous pragmatic influences – colour “hot spots” in the speakers’ environment specific to the language/culture – that affect partitioning of the original composite category. The “hot spots” serve as semantically shifted “anchors” in the formation of the incipient colour category, stipulating its high naming consensus in the population and elevating it to basic status. Conversely, in this “handicapped partitioning”, the areas of low intra-population consensus within the composite category shrink and fall out of basicness (Gooyabadi et al. 2019: 169). The computer-simulation findings are in accord with MacLaury’s (2007: 145) observations in the Mesoamerican Color Survey: “Many desaturated-complex categories are curtailed in range by addition to them of emphasis on either light or dark or on hue ... they are replaced by a simpler alternative that comes complete with a new name ... [of] the simpler construct.”

One needs to remark that the empirical World Color Survey and Mesoamerican Color Survey studies, as well as their computer simulations investigated the evolutionary dynamics of BCCs/BCTs using a cross-language analysis, so can be considered only as a proxy of diachronic analysis of an individual language. The failure of the optimality principle of the colour gamut’s lexical refinement in languages with complex colour inventories assumed in those studies (Jraissati & Douven 2017) points to other factors, such as interaction between cultural needs and environmental factors, intercultural exchange, or historical heritage that are important for understanding the evolution of colour concepts in a diachronic perspective (Decock 2021; Zaslavsky et al. 2022).

References

Decock, Lieven. 2021. Conceptual change and conceptual engineering: the case of colour concepts. *Inquiry* 64(1-2). 168–185. <https://doi.org/10.1080/0020174X.2020.1784783>

- Gooyabadi, Maryam, Kirbi Joe, and Louis Narens. 2019. Further evolution of natural categorization systems: An approach to evolving color concepts. *Journal of the Optical Society of America A* 36(2). 159–172. <https://doi.org/10.1364/JOSAA.36.000159>
- Jraissati, Yasmina and Igor Douven. 2017. Does optimal partitioning of color space account for universal color categorization? *PLoS ONE* 12(6). e0178083. <https://doi.org/10.1371/journal.pone.0178083>
- Komarova, Natalia L., Kimberly A. Jameson, and Louis Narens. 2007. Evolutionary models of color categorization based on discrimination. *Journal of Mathematical Psychology* 51(6). 359–382. <https://doi.org/10.1016/j.jmp.2007.06.001>
- MacLaury, Robert E. 2007. Categories of desaturated-complex color. Sensorial, perceptual, and cognitive models. In Robert E. MacLaury, Galina V. Paramei, and Don Dedrick (eds.), *Anthropology of color: Interdisciplinary multilevel modeling*, 125–150. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/z.137.10mac>
- Steels, Luc. 2011. Modeling the cultural evolution of language. *Physics of Life Reviews* 8(4). 339–356. <https://doi.org/10.1016/j.plrev.2011.10.014>
- Zaslavsky, Noga, Karee Garvin, Charles Kemp, Naftali Tishby, and Terry Regier. 2022. The evolution of color naming reflects pressure for efficiency: Evidence from the recent past. *Journal of Language Evolution*. Early view. lzac001. <https://doi.org/10.1093/jole/lzac001>

Supplement S3. History of the colour term *buryj* (Bakhilina 1975: 227–228)

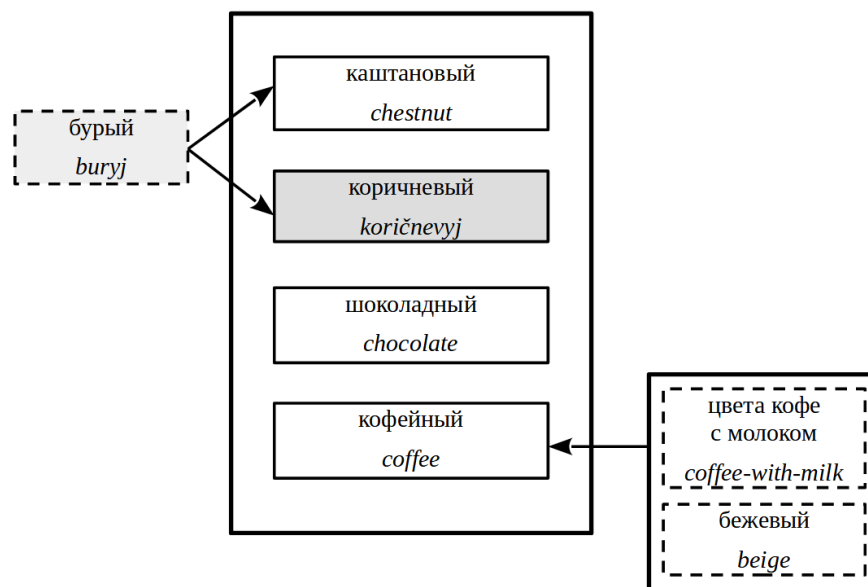
“The history of the adjective *buryj* as a colour term conceivably unfolded as follows. The adjective *buryj* was borrowed in old times, was broadly used for a very long time, and seemingly already in old times had been a colour term with almost unconstrained combinability. It is this term that in the folk language had always been used for denoting shades of the colour brown. At the same time, as a colour term, it seems to have been rather imprecise, and named very different shades of brown. With time, the scope of word use had not narrowed but during some historical periods, and within some usage contexts, the word acquired an element of emotional-expressive appraisal. This is an undefined, “not right” and unpleasant colour. It may be that the emotional-expressive connotation hampered the word from becoming a neutral denotation of the brown colour. It is quite probable that its fate had been affected by the emergence in the language of the word *koričnevyyj*, whose neutral meaning gave it greater potential to become an abstract colour term.” (Translation of the original Russian text by one of the authors, GVP.)

References

Bakhilina, Natalia B. 1975. *Istorija cvetooboznačenij v russkom jazyke [History of colour terms in Russian]*. Moscow: Nauka (in Russian).

Supplement 4: Semantic cluster of most frequent terms for 'brown' in modern Russian

In modern Russian, *buryj* is semantically peripheral to the 'brown'-cluster derived from a psycholinguistic analysis (based on free-sorting of Munsell chips), the cluster dominated by *koričnevuj* (Frumkina 1984: 68).



The structure of the cluster of Russian 'brown' colour terms. Adapted from Frumkina (1984: 68).

References

Frumkina, Rebecca M. 1984. *Cvet, smysl, srodstvo. Aspekty psixolingvističeskogo analiza [Colour, meaning, and similarity: Aspects of a psycholinguistic analysis]*. Moscow: Nauka (in Russian).

**Supplement S5: Historical details of the introduction of cinnamon in Russia and
koričnevij entrenching**

The term *korica* is first recorded in 1472 in the book *Voyage Beyond Three Seas 1466–1472* by Afanasy Nikitin, a Russian merchant and one of the first Europeans to travel to Persia and India. The Russian term for cinnamon, *korica*, appears to be a neologism created by Nikitin as a diminutive of Russian *кора / kora* ‘bark’. Originally, to ensure comprehension, he accompanied the term *korica* by the Cyrillic transliteration of *cinamomum*, the Indian name for the Ceylon cinnamon tree (<https://www.nkj.ru/archive/articles/13293/>).

Before Nikitin’s voyage, cinnamon for centuries had been of considerable trading importance in Asia (e.g. Suriyagoda et al. 2021). It was listed among the trade “marvels” – as a precious spice used in cooking and, also, highly valued for its medicinal bioactive properties and health benefits against common diseases and disorders.

Cinnamon rapidly acquired socio-cultural significance in Russia: in the 16th and 17th centuries, there are records of deliveries of significant amounts of cinnamon to Russia (by Persian merchants) since “it was very much liked by Russians” (Spassky 1910: 49). As an essential ingredient of the Russian traditional *sbiten’* (a hot mead drink), cinnamon is mentioned in *Domostroy*, the 16th-century Russian management manual that contained a fundamental set of household rules, instructions and advice. In the 16th- and 17th-century texts there are also abundant recipes for dishes (poultry, meat), food products (vodka, butter), and medicines (including aphrodisiacs) containing cinnamon (Slovari 11–17 vekov 1975–: 314).

From a diachronic perspective, an emerging CT enters Russian as a denominal adjective *X-yj* and is established gradually. Along with (a) the “colour” meaning, it may also possess the meanings such as (b) “made of X/containing X” or (c) “related to X” (Rakhilina 2007a). In the 16th–18th century dictionaries, one finds various adjectival forms of *korica* that develop either individual meanings (a-c) or combinations of these. With the meaning ‘colour of *korica*’, the term *koričnevij* (also spelled *korišnevij*, *korišnyj*, *korišnevij* or *koričnyj*) is recorded for the first time in 17th-century texts denoting the brown colour of textiles (Bakhilina 1975: 228). In comparison, the form *koričnyj/ korišnyj* collocates with ‘vodka’ and ‘butter’, indicating products made of cinnamon, i.e. the meaning (b). Finally, in relation to the bark of the cinnamon tree (c), several adjectival forms are recorded – *koričnevij*,

korišnevyj, *korišnyj* and *koričij* (Slovar' Akademii Rossijskoj 1789–1794, vol. 3: 821; Slovarei 11–17 vekov 1975–: 358).

The socio-cultural (“culinary”) circumstances – fondness of cinnamon – made Russian vocabulary conducive to the lexical refinement of the BROWN category by following the cross-language pattern of colour-name metonymy (Murjanov 1978; Biggam 2012): the term for the object “which instantiates the best example of colour also represents that colour” (Steinvall 2002: 143). In other languages, similar ‘referent object’ cases of ‘brown’ lexicalisation are exemplified by Maltese *kannella* (from Italian *cannella* ‘cinnamon’; Borg 2011); *marron* in French (Forbes 1979) and *marrón* in Castilian Spanish (Lillo et al. 2018); or *café* in Mexican Spanish (Lillo et al. 2018) and *kāfēi* in modern Taiwanese Mandarin Chinese (Hsieh et al. 2020). Notably, Russian lexicalisation of the new ‘brown’ term took on an alternative way of coining, as pointed out by Kerttula (2002): if the meaning of the borrowed term is unknown to speakers of the receiving language, it is often the term’s translation which is adopted.

References

- Bakhilina, Natalia B. 1975. *Istorija cvetooboznačenij v russkom jazyke [History of colour terms in Russian]*. Moscow: Nauka (in Russian).
- Biggam, Carole P. 2012. *The semantics of colour: A historical approach*. Cambridge, UK: Cambridge University Press.
- Borg, Alexander. 2011. Towards a diachrony of Maltese basic colour terms. In Carole P. Biggam, Carole A. Hough, Christian J. Kay, & David R. Simmons (eds.), *New directions in colour studies*, 73–90). Amsterdam/Philadelphia: John Benjamins.
- Forbes, Isabel. 1979. The terms *brun* and *marron* in modern standard French. *Journal of Linguistics* 15(2). 295–305. <https://www.jstor.org/stable/4175499>
- Hsieh, Tsuei-Ju T., Ichiro Kuriki, I-Ping Chen, Yumiko Muto et al. 2020. Basic color categories in Mandarin Chinese revealed by cluster analysis. *Journal of Vision* 20(12). 6. <https://doi.org/10.1167/jov.20.12.6>
- Kerttula, Seija. 2002. *English colour terms: Etymology, chronology, and relative basicness*. *Mémoires de la Société Néophilologique de Helsinki*, Vol. LX. Helsinki: Société Néophilologique.
- Lillo, Julio, Lilia Prado-León, Fernando G. Perilli, Anna Melnikova et al. 2018. Spanish basic colour categories are 11 or 12 depending on the dialect. In Lindsay W. MacDonald, Carole P. Biggam, and Galina V. Paramei (eds.), *Progress in colour studies: Cognition, language and beyond*, 59–82. Amsterdam/Philadelphia: John Benjamins.
- Murjanov, Mikhail F. (1978). K interpretacii sratoslavjanskix tsvetooboznačenij [On the interpretation of Old Church Slavonic colour terms]. *Voprosy jazykoznanija* Iss. 5. 93–109 (in Russian).

- Rakhilina, Ekaterina V. 2007. Linguistic construal of colors: The case of Russian. In Robert E. MacLaury, Galina V. Paramei, and Don Dedrick (eds.), *Anthropology of color: Interdisciplinary multilevel modeling*, 363–377. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/z.137.24rak>
- Slovari russkogo jazyka 11–17 vekov: Ėtimologija i istorija slov russkogo jazyka* [Dictionaries of the Russian language of the 11th–17th centuries: Etymology and history of Russian words] (1975–) (pp. 314, 358). Moscow: Nauka (in Russian).
- Slovar' Akademii Rossijskoj* [Dictionary of the Russian Academy]. 1789–1794. St. Petersburg: Imperial Academy of Sciences (in Russian).
- Spassky, Pavel K. (ed.). 1910. *Istorija trgovli i promyšlennosti v Rossii* [The history of trade and industry in Russia], Vol. 1. St. Petersburg (in Russian).
- Steinvall, Anders. 2002. *English colour terms in context*. PhD thesis, Umeå University.
- Suriyagoda, Lalith, Anoma J. Mohotti, Janak K. Vidanarachchi, Suranga P. Kodithuwakku et al. 2021. “Ceylon cinnamon”: Much more than just a spice. *Plants, People, Planet* 3(4). 319–336. <https://doi.org/10.1002/ppp3.10192>

Table S1. Dynamics of the derivational productivity of *buryj* and *koričnyj*: Comparison of derivational forms reported by Corbett and Morgan (1988: 38) and recorded in the present study. Novel recorded forms are in red: Russian subcorpus of Google Books Ngram (1800–2019) in bold; occurrences on the Internet and recent dictionaries are in plain font.

Derivational forms	Source	Suffixed	Achromatic modifiers					Compound	Verbs			Noun
		-ovatyj '-ish'	tëmno- 'dark'	svetlo- 'light'	jarko- 'bright'	bledno- 'pale'	blëklo- 'pallid'	sero- 'grey'	∅~et' '-en'	po~et' 'become ~'	za~et' 'to acquire ~ shade'	-izna '-ness'
<i>buryj</i> 'dust/greyish brown'	Corbett & Morgan (1988)	√	√					√	√	√	√	
	Present study	√	√	√	√	√	√	√	√	√	√	√
<i>koričnyj</i> 'brown'	Corbett & Morgan (1988)	√		√								
	Present study	√	√	√	√	√	√	√	√	√	√	√

Corbett and Morgan's (1988) data were collated based on the report of Worth et al. (1970) who, in turn, leaned upon Ozhegov and Shapiro's (1959), then the most comprehensive dictionary of Russian. The latter had been prepared over several years preceding its publication and based on early dictionary materials; therefore, it cannot fully reflect the state of the Russian language post-1956, when the Russian language orthographic reform was introduced.

References

- Corbett, Greville G. and Gerry Morgan. 1988. Colour terms in Russian: Reflections of typological constraints in a single language. *Journal of Linguistics* 24(1). 31–64. <https://www.jstor.org/stable/4175920>
- Ozhegov, Sergei I. and Abram B. Shapiro. (eds.). 1959. *Orfografičeskij slovar' russkogo jazyka* [Orthographic dictionary of the Russian language], 4th ed. Moscow: State publisher of foreign and national dictionaries (in Russian).
- Worth, Dean S., Andrew S., Kozak, and Donald B. Johnson. 1970. *Russian derivational dictionary*. New York: Elsevier.

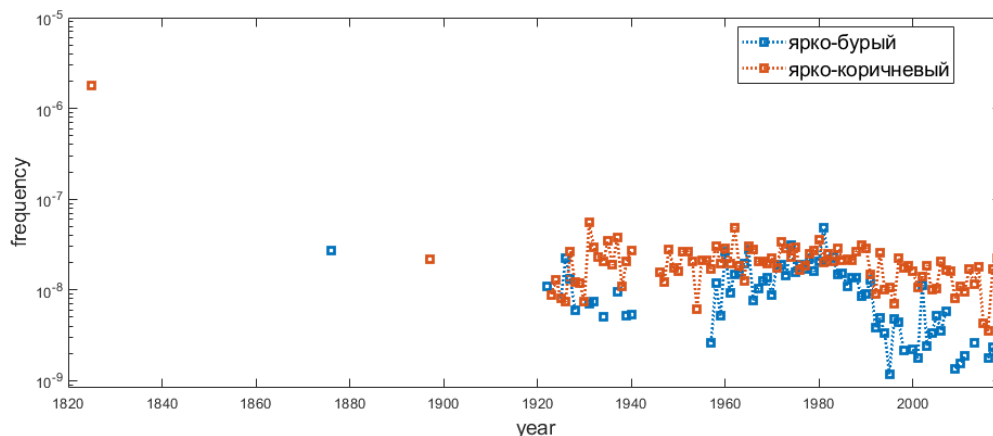


Figure S1. Diachronic dynamics of the two 'Russian browns' with the modifier *jarko-* 'bright': *ярко-бурый* 'bright *buryj*' (—) and *ярко-коричневый* 'bright *korišnevyj*' (—).

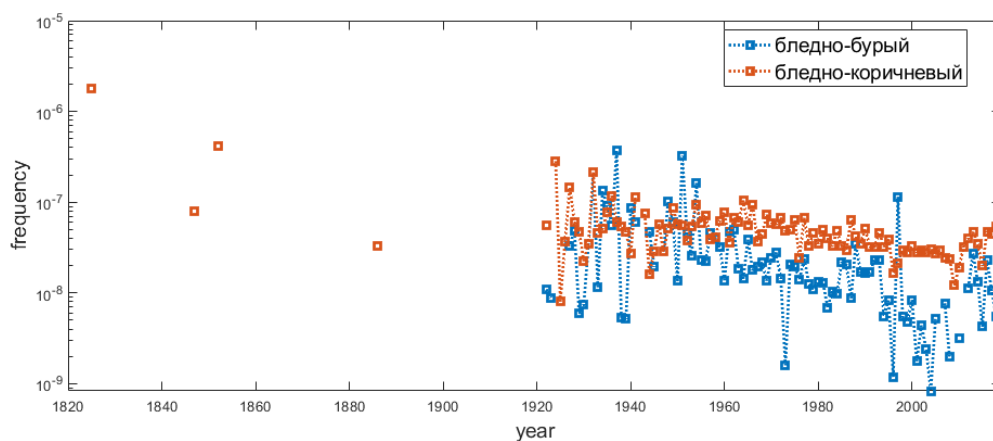


Figure S2. Diachronic dynamics of the two 'Russian browns' with modifier *bledno-* 'pale': *бледно-бурый* 'pale *buryj*' (—) and *бледно-коричневый* 'pale *korišnevyj*' (—).

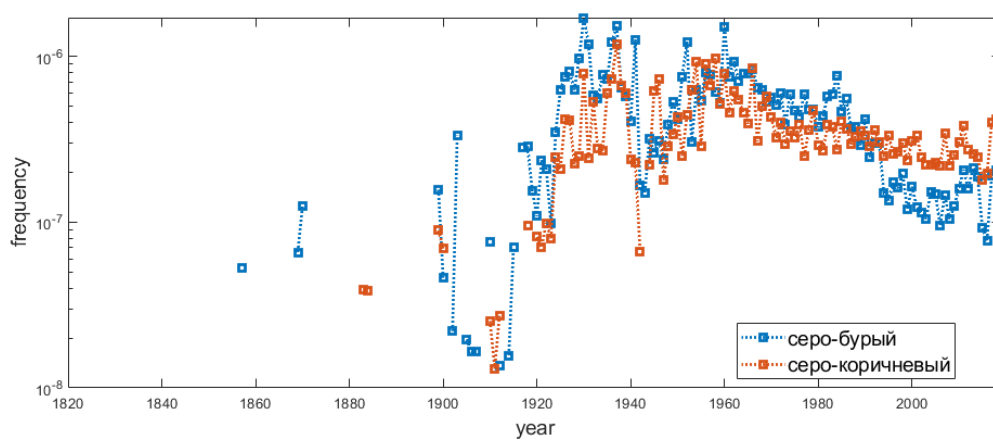


Figure S3. Diachronic dynamics of the two 'Russian browns' with the compound *sero-* 'grey': *серо-бурый* 'grey-*buryj*' (—) and *серо-коричневый* 'grey-*korišnevyj*' (—).

Supplementary Examples (SE) of *buryj*- and *koričnevuj*-collocations, and their verbal and noun derivatives in the Russian National Corpus (RNC) and Internet sources

SE1. Recent instances of *blëklo*- ‘pallid’ modifier with *buryj* and *koričnevuj*.

“... в эксперименте на расчищенной от травы площадке **блëкло-бурого** цвета.”

Сонин, Н., Агафонова, И., Захаров, В. (2014). Биология. Общие закономерности. 9 класс. 5-е изд. Москва: Дрофа.

“Мякоть **блëкло-коричневая**, тонкая, рассыпчатая, с пряным запахом.” Матанцев, А., Матанцева, С. (2017). Большая энциклопедия грибника. Москва: АСТ.

SE2. Instances of **inceptive prefixed verbal derivative** of *buryj* – *zaburet* ‘to become *buryj*’:

“Наступило самое жаркое время; начался покос, рожь **забурела**; знойный, удушливый ветер лениво бродил по озёрам, чуть-чуть нагибая верхи камышей.” Слепцов, В.А. (1865). Трудное время (cit. from the RNC).

“На некрасивом, суровом лице есаула Афанасия Перфильева **забурели** старые следы ослы.” Шишков, В.Я. (1934–1939). Емельян Пугачёв. Книга первая, Ч. 3. (cit. from the RNC).

Note also an additional novel semantics of *zaburet* in younger generations’ vernacular, meaning ‘to become impudent, overestimate oneself’: see Зайковская, Т.В. (2005).

Проблемы культуры молодежной речи: пути пополнения словарного состава молодежного жаргона. Кишинев: Busines-Elita.

GBN records this verbal form in both meanings.

SE3. Recent Internet-only instances of **inceptive prefixed verbal derivatives** of *koričnevuj*:

zakoričnevet ‘to become *koričnevuj*’

“Растопить сливочное масло и дать ему слегка **закоричневеть**.”

(<https://www.oede.ru/component/cockbook/dish/1613-dish.html>)

pokoričnevet ‘to gain *koričnevuj* taint’

“Их деревья когда-то были зелёными, но затем с ними произошло то же, что ежегодно происходит с реальными деревьями в лесу. Они пожелтели, **покоричневели**” .Цукерман, В. (1970). Все краски мира. Химия и жизнь. No. 6, с. 41–47.

“Чтобы слегка **покоричневеть**, нужно не покидать пляж с утра до вечера и так целый месяц.” (https://bookz.ru/authors/nadejda-volgina/vendetta_741/1-vendetta_741.html)

SE4. GBN and Internet instances of **nouns X-izna** ‘X-ness’ derived from *buryj* and *koričnevuj*.

“болѣ слабомѣ просвѣчиваниі бурыхъ пятнышекъ замѣчается пятнистая **буризна**; ...такихъ болѣ крупныхъ бурыхъ пятенъ насчитывается шесть...” [Монголія и Камь](#).

[Труды экспедиции Императорского Русского Географического общества. \(1908\). Том.7. вып. 1.](#)

“Из-за этих продуктов действительно возможна некая **буризна** и краснота.” 2018. (<https://lapkins.ru/f/korm-dlya-belykh-sobak-163/>)

“У этой конктерной лошади заметная **коричневизна** на боку, у булано-саврасых такого не бывает.” (<http://raiter.flyboard.ru/topic17-945.html?view=print>)

Examples of atypical *koričnevuj*-collocations with nouns denoting natural objects

Медведь *medved'* ‘bear’ / медвежонок *medvežonok* ‘bear cub’:

SE5. Descriptions in paintings: “Грузный **коричневый медведь** дерется с разъяренным быком, рядом, потехи ради, карлик, преследуемый кабаном.” [Каптерева, Т. \(2017\). Римская мозаика. Африка. Россия: ЛитРес, с. 32.](#)

SE6. References to toys (teddy bears) and souvenirs: “Они проходили мимо стрелкового тира, где на витрине красовался большой **коричневый медведь** с атласной красной ленточкой на шее.” [Лонго, С. \(2018\). Арабеска зеркал. Россия: ЛитРес.](#)

SE7. A character in Russian fairy-tales: “На пороге стоял **коричневый медвежонок** в зелёной пижаме”. [Танилина, Т. \(2017\). Мяуняш. Россия: ЛитРес.](#)

Russian translations by *koričnevuj* of English *brown* in collocations with nouns denoting other animals (fiction):

SE8. (snake) “Откуда-то из-за кустов выползла **коричневая** змея, но Мэтт отогнал ее.” [Макинтош, Ф. \(2022\). Возвращение в Прованс. Россия: Эксмо.](#) (McIntosh, F. The French promise.)

SE9. (spider) “Когда я вернулась в свою хижину, то обнаружила, что забыла закрепить москитную сетку вокруг кровати, и в постель пробрался **коричневый паук** с толстыми, покрытыми наростами лапами.” [Филдинг, Х. \(2019\). Причина успеха. Россия: Амфора.](#) (Fielding, H. Cause celeb.)

Russian translations of English *brown* by *koričnevuj* in zoological terms:

SE10. (snake) *Australian eastern brown snake* → *австралийская восточная коричневая змея*. [Ларин-Подольский, И. \(2017\). Юбилейные и памятные монеты мира. Иллюстрированная энциклопедия. Россия: ЛитРес, с. 233.](#)

SE11. (spider) “В Соединенных Штатах есть три вида пауков, которые опасны для человека: **коричневый** паук-отшельник, американский бродячий паук и черная вдова.” [Хаббард, Д. \(2017\). Первая помощь своими руками: Если скорая не спешит. \(п.р.\): Альпина Паблишер](#) (Hubbard, J. The survival doctor’s complete handbook: What to do when help is NOT on the way.)

Koričnevuj characters in modern Russian fairy-tales:

SE12. (spider) “В самом центре восседал президент Пантелеймон – старый, волосатый, **светло-коричневый** паук с белым крестом на спине.” [Трушкина, О. \(2018\). Странствия Игорца. Россия: ЛитРес.](#)

In Russian fiction, part of a metaphor alluding to an artefact taking a snake-like shape (e.g. a scarf):

SE13. “...**шарф**, который , как шерстяная коричневая змея, притаился в изножье кровати.” [Кёртин, Д. \(2018\). Добро пожаловать во Францию, Элис!. Россия: Эксмо.](#)

Koričnevuj in Russian (translations of) ‘brown’ characterising processed minerals:

SE14. (marble): “Полы в приемной были выложены **коричневым** мрамором, а ресепшен представлял собой архитектурное подобие площади какого-то античного города.” [Фарутин, А. \(2019\). Карьерист. Трилогия. Россия: ЛитРес.](#)

SE15. (diamond) “Хуан вспомнил **светло-коричневый** алмаз из серьги королевы и задумался.” [Боуэн, М. \(2019\). Рыцарь Испании. Россия: ЛитРес.](#) (Bowen, M. A Knight of Spain.)

Koričnevuj as the Russian term for brown sugar:

SE16. “В жестянке вареный **коричневый** сахар, порубленный на кусочки, – конфеты домашнего приготовления, называемые вдовой «крем брюле».” [Шагинян, М.С. \(1941\). Два романа. Россия: Сов. писатель.](#)

Koričnevuj as a part of metonymies implying concepts of evil

SE17. (brown dope, spell) **коричневый дурман** (“О злодеяниях немецко-фашистских оккупантов в Ставропольском крае” и статье А.Н. Толстого “Коричневый дурман”) [Перхин, В. \(2018\). А.Н. Толстой и власть. Россия: ЛитРес.](#)

SE18. (brown scum) “Тем временем другая, прожорливая и кровожадная, **коричневая нечисть**, расплзаясь по Европе,” [Прасолов, В. \(2018\). Вангол. Россия: Центрпол.](#)

SE19. (brown army) “По **коричневой армии** пронесся лозунг «Третьей Империи», полный магического, едва ли не религиозного энтузиазма”. [Устрялов, Н. \(2017\). В круговороте фашистской свастики. Россия: Алгоритм.](#)

SE20. (brown plague) “Да, и красная, и **коричневая чума** в значительной степени исчезли с лица земли, но успели унести миллионы жизней!” [Эпштейн, М. \(2019\). Постмодернизм в России. Россия: Азбука-Аттикус.](#)

Instances of *buryj*-collocations with nouns for artefacts, in which a drab brown is implied to convey a negative connotation

SE21. (brown factory chimneys) “Две **бурые трубы** завода непрерывно источали копоть.” [Дмитриев, В., Богданов, Н. \(1929\). Александр и Александра. Смена. No. 123.](#)

SE22. [(bloodshot) brown neck; brown spots (stumps of fingers)] “...тесный воротник его старенькой толстовки, врезавшийся в **бурую**, туго налитую **шею**...”, “...на изуродованной правой руке Устина одиноко торчит указательный палец, а на месте остальных темнеют **бурые**, сморщенные **пятна**.” [Шолохов, М.А. \(1986\). Поднятая целина. Минск: Полымя.](#)

SE23. [(worn out) brown blouse] “Федя косился на свою **бурую блузу**, на рыжие, в трещинах и порезах, сапоги...” [Ляшко, Н.Н. \(1955\). Сочинения в трех томах. т. 2, Россия: Гос. изд-во худ. лит-ры, с. 398.](#)

SE24. [(arthritic) brown fingers] “Узловатые, **бурые пальцы** вытянутой вперед руки нервно изгибались дождевыми червями.” [Серебрякова, Г. \(1933\). Юность Маркса. Новый мир, No. 4, с. 155.](#)