LINGUISTIK INTERNATIONAL

HERAUSGEGEBEN VON

HEINRICH WEBER
WILFRIED KÜRSCHNER
REINHARD RAPP
KAZIMIERZ SROKA
JÜRG STRÄSSLER
LEW ZYBATOW

BAND 44

PETER LANG

LUKA SZUCSICH / AGNES KIM / ULIANA YAZHINOVA
(EDS.)

AREAL CONVERGENCE IN EASTERN CENTRAL EUROPEAN LANGUAGES AND BEYOND

PETER LANG

This publication has been peer reviewed.

www.peterlang.com

Contents

List of Contributors

Agnes Kim
University of Vienna

Anna Tetereva
Humboldt University Berlin

Emmerich Kelih
University of Vienna

František Martínek
Charles University/Prague

Ivan Šimko
University of Vienna

Jerzy Gaszewski
(ORCID 0000-0002-9313-6454)
University of Wrocław

Jiří Januška
Charles University/Prague

Luka Szucsich
Humboldt University Berlin

Nataliya Levkovych
University of Bremen

Sebastian Scharf
Humboldt University Berlin

Stefan Michael Newerkla
University of Vienna

Tamás Tölgyesi
University of Vienna

Thomas Stolz
University of Bremen

Uliana Yazhinova
Humboldt University Berlin

Viktoria Naukhatskaia
Humboldt University Berlin

RV 217 AB 299 S. 30. BR: AB 1557 S. 354.) [= Federal Act dated 7th July 1976 on the legal status of ethnic groups in Austria].

Wistrich, Robert S.: *Laboratory for world destruction: Germans and Jews in Central Europe*. University of Nebraska Press: Lincoln, NE. 2007.

Wittmann, Henri: Classification linguistique des langues signées non vocalement. In: *Revue québécoise de linguistique théorique et appliquée* 10(1), 1991, pp. 215–288.

Zeman, Dalibor: *Überlegungen zur deutschen Sprache in Österreich: linguistische, sprachpolitische und soziolinguistische Aspekte der österreichischen Varietät*. Kovac, Dr. Verlag: Hamburg. 2009.

Ivan Šimko & Emmerich Kelih

Loanwords in Bulgarian Core Vocabulary – a Pilot Study

Abstract: The following text presents methodological reflections on the pilot study of loanwords in South Slavic languages. The study focuses on Bulgarian core vocabulary, using the word list by Carlton (1990) as a reference corpus. The list includes words of various semantic and grammatical categories, which are considered relatively stable, old and resistant to borrowings (e.g. Swadesh 1952). The authors use a method which marks the status of a word as a likely borrowing from the World Loanword Database (Haspelmath/Tadmor 2009b) and adapt it to criteria specific for Slavic etymological studies.

Keywords: loanwords, Bulgarian, language contact, language borders, etymology

1 Introduction

One of the most obvious results of language contact is the borrowing of lexical items. The analysis of lexical borrowings has a long tradition in Slavic studies, as they present a valuable source of information about the historical development of both the language and those who speak it. As relics of past contact situations, loanwords often reflect both cultural and natural phenomena which were previously unknown to the language community. The community, emerging from its previous isolation, expanded its vocabulary and thus its horizons as well. In contrast to this layer of newly acquired concepts, stands the idea of the core vocabulary of the language: words for everyday phenomena, where we can only imagine an extralinguistic motive for a borrowing. The question is how to determine the frequency of loanwords within this layer of vocabulary. What does their presence tell us about the intensity of contact between the donor and recipient languages? Within our project of South Slavic loanword studies, we have already analysed the Slovene (Kelih 2015), Croatian (Kelih/Garić 2016) and Bulgarian (Kelih/Šimko 2018) core vocabularies for loanwords. In this study we will present some of the methodological questions encountered by the authors during their analysis of the Bulgarian language.

2 Borrowings in the Core Vocabulary

The idea of a particular lexical layer which is particularly resistant to borrowing is based on the fact that we can already observe a transfer of lexical units in

(very) casual language contact situations. The most obvious case is the adoption of a lexeme for a concrete noun which was not present in the language before, e.g. words describing new technologies or economic relations. Only if the contact is intense, can we also observe borrowings of abstract nouns, as well as concrete ones for concepts already present in the recipient language. In a number of studies (e.g. Swadesh 1952, Embleton 1986, D'Andrade 1995, Hock/Joseph 1996, p. 257, Zenner/Speelman/Geeraerts 2014) it is emphasised that the core vocabulary of a language (according to Swadesh, the founder of glottochronology, some kind of culture-free list of lexical items) is a relatively old, stable lexical stratum, almost resistant to borrowings. It is therefore not supposed to be subject to greater changes.

However, contrastive studies of various languages (Kelih 2015, Tadmor/Haspelmath/Taylor 2010, Haspelmath 2009b, Haarmann 1990) have shown that in fact the core vocabulary of a language also integrates borrowings to a certain degree. Bulgarian has had a word *pivo* for 'beer' at least since the modern period, but it is being replaced by the Italian loanword *bira*. From the historical view, such borrowings are not always simple replacements. For example, the word *hora* 'people' was originally borrowed from Greek in its original meaning 'land'. Later it was used for the 'inhabitants of the land', and finally as 'people' in general, replacing the Slavic root *ljude*.

Moreover—and this makes the analysis of loanwords in the core vocabulary linguistically interesting—the amount of borrowing in the core vocabulary varies depending on the intensity of language contact, which has an impact on the semantic fields of the items being borrowed. An important recent resource for the amount and kind of lexical borrowing can be found in the World Loanword Database[1] (WOLD, Haspelmath/Tadmor 2009b), which provides an overview of borrowings in the core vocabularies (containing 1,500 lexical meanings in 24 different semantic fields) of over 40 languages, with different numbers of speakers, historical contexts and sociolinguistic statuses. The results of this crosslinguistic study (Tadmor 2009) clearly show both a language-specific incorporation rate and an individual distribution of loanwords among particular semantic fields (e.g. religion, clothing, home, kinship terms, emotions etc.). Based on these findings it appears that the core vocabulary of a language has to be understood as the result of various impact factors like the depth and intensity of language contact situations, puristic attitudes and their particular influence on

the standardisation process, the extent of multilingualism in a language community, structural incompatibility, genealogical relatedness and many others.

Coming back to the initial question it has to be mentioned that according to our knowledge there is no comparable systematic study of borrowings in the core vocabulary of South Slavic languages; the only representative of the Slavic family in the WOLD project is Lower Sorbian. To give a general idea of the analysis of loanwords in the core vocabulary for the purposes of our pilot study of the Bulgarian core vocabulary we used a modified version of Swadesh's well-known basic vocabulary list for Bulgarian, compiled by Carlton (cf. 1990, pp. 334–349) in his book on Slavic historical phonology.

The original Swadesh list was extended to 212 entries (the original had 200), grouped into eight semantic and grammatical categories: (1) common adjectives, (2) common animals and birds, (3) common plants, (4) common verbs, (5) kinship terms, (6) a group of concepts concerning nature, tools and housing, (7) concepts concerning nourishment, and (8) body parts. The list includes the vocabulary in 12 standard Slavic languages, two attested older languages (Old Church Slavonic and Polabian), as well as in the reconstructed Proto-Slavic. As the list is based on an earlier work by Meľnyčuk (1966), in the following we refer to this list as "SMC" (Swadesh–Meľnyčuk–Carlton). For a better overview, we add meanings of the respective words in modern Bulgarian:

(1) *bjal* 'white', *cjal* 'whole', *čist* 'clean', *čeren* 'black', *červen* 'red', *dălăg* 'long', *dobăr* 'good', *gorăk* 'bitter', *kisel* 'sour', *krasen* 'beautiful', *libe* 'beloved' (actually a substantive), *malăk* 'small', *mek* 'soft', *mlad* 'young', *pălen* 'full', *zdrav* 'healthy', *slab* 'weak', *star* 'old', *čužd* 'strange, foreign', *velik* 'great', *zelen* 'green', *zăl* 'angry', *živ* 'alive', *žălt* 'yellow'.

(2) *agne* 'lamb', *jajce* 'egg', *bobăr* 'beaver', *pčela* 'bee', *bik* 'bull', *červej* 'worm', *elen* 'deer', *esetra* 'sturgeon', *găska* 'goose', *ež* 'hedgehog', *kobila* 'mare', *kon* 'horse', *krava* 'cow', *koza* 'goat', *kur* 'penis' (earlier 'rooster'), *mravka* 'ant', *lebed* 'swan', *orel* 'eagle', *osa* 'wasp', *ovca* 'sheep', *pes* 'dog', *prase* 'pig', *riba* 'fish', *skot* 'cattle', *svinja* 'sow', *svraka* 'magpie', *tele* 'calf', *tur* 'aurochs', *vepăr* 'swine', *vălk* 'wolf', *vol* 'ox', *vrana* 'crow', *žrebec* 'stallion', *zmija* 'snake', *zvjar* 'wild animal'.

(3) *jabălka* 'apple', *jagoda* 'strawberry', *breza* 'birch', *bob* 'bean', *buk* 'beech', *băz* 'elder', *čereša* 'cherry', *česăn* 'garlic', *dărvo* 'tree', *dăb* 'oak', *ečmik* 'barley', *ela* 'fir', *elha* 'alder', *gabăr* 'hornbeam', *kruša* 'pear', *klen* 'maple', *kopăr* 'dill', *cvjat* 'flower', *lipa* 'lime tree', *len* 'flax', *malina* 'raspberry', *oreh* 'nut', *oves* 'oat', *proso* 'millet', *pšenica* 'wheat', *răž* 'rye', *sliva* 'plum', *smreka* 'juniper, spruce', *treva* 'grass', *vărba* 'willow', *želăd* 'acorn', *žito* 'grain'.

(4) *bjagam* 'run', *bija* 'beat', *boli* 'hurt', *săm* 'be', *češa* 'comb', *čuja* 'hear', *dam* 'give', *dărža* 'hold', *jaham* 'ride', *šta* 'want', *ida* 'go', *kălna* 'swear', *kărmja* 'nurse, feed', *melja* 'mill, grind', *molja* 'please, pray', *mălča* 'be silent', *peka* 'bake', *seja* 'sow', *tresa* 'tremble', *tăka* 'weave', *varja* 'boil', *vozja* 'carry', *veža* 'bind', *živeja* 'live'.

---

1   Cf. http://wold.clld.org (11.06.2017.)

(5) *baba* 'grandmother', *brat* 'brother', *čovek* 'human', *djado* 'grandfather', *dete* 'child', *dever* 'brother-in-law', *dăšterja* 'daughter', *ljude* 'people', *măž* 'man, husband', *nevesta* 'bride, daughter-in-law', *otec* 'father', *sestra* 'sister', *sin* 'son', *svekăr* 'father-in-law (bridegroom's father)', *tăst* 'father-in-law (bride's father)', *vnuk* 'grandchild', *zet* 'son-in-law', *žena* 'woman, spouse'.

(6) *brjag* 'coast, shore', *brana* 'harrow', *cep* 'chain', *čad* 'haze', *člun* 'boat', *den* 'day', *dol* 'valley', *dom* 'home', *dăžd* 'rain', *dim* 'smoke', *dveri* 'door', *dvor* 'court, yard', *ezero* 'lake', *gora* 'forest' (earlier 'mountain'), *zvezda* 'star', *kamen* 'stone', *ljato* 'summer', *mesec* 'moon, month', *nošt* 'night', *ogăn* 'fire', *os* 'axle', *plug* 'plough', *pole* 'field', *reka* 'river', *rosa* 'dew', *snjag* 'snow', *slănce* 'sun', *vjatăr* 'wind', *voda* 'water', *voz* 'cart load' (earlier 'cart, wagon'), *zemja* 'earth', *zlato* 'gold'.

(7) *doja* 'milk', *jam* 'eat', *hljab* 'bread', *kvas* 'yeast', *loj* 'tallow', *maslo* 'fat, grease', *med* 'honey', *mljako* 'milk', *meso* 'meat', *pija* 'drink', *pivo* 'beer', *salo* 'fat', *sirene* '(white) cheese', *sit* 'fed', *testo* 'dough', *vino* 'wine'.

(8) *brada* 'beard', *čelo* 'forehead', *čeljust* 'jaw', *červo* 'gut', *dlan* 'palm', *glava* 'head', *ezik* 'tongue', *kost* 'bone', *koža* 'skin', *krăv* 'blood', *noga* 'leg', *nokăt* 'nail', *nos* 'nose', *oko* 'eye', *lakăt* 'elbow', *palec* 'thumb', *peta* 'heel', *prăst* 'finger', *răka* 'hand', *sărce* 'heart', *tjalo* 'body', *vălna* 'wave', *vime* 'udder', *zăb* 'tooth'.

The identification and determination of loanwords in the SMC list is accompanied by several linguistic problems, namely:

1. The list doesn't include the meanings of the particular lexemes, which are only grouped roughly into semantic fields. Although some words like *dom* 'home' or *voz* 'cart load' are still attestable in modern Bulgarian, their meaning is different from their cognates in other Slavic languages, like Old Church Slavonic *domъ* 'house', *vozъ* 'cart' (cf. Cejtlin 1994, p. 194; Ilčev 1998, p. 66; *Rečnik* I, p. 171). For the concepts of 'house' and 'cart' modern Bulgarian uses the words *kăšta* and *karuca* – in the latter case, a clear borrowing (cf. Ilčev 1998, p. 66; *Rečnik* II, p. 256).

2. Some of the chosen semantic categories themselves lack clear boundaries: e.g. Bulgarian lacks a reflex for **lěsъ*, grouped under "Common plants", but in fact standard Bulgarian has replaced this old word[2] for 'forest' with *gora*, which can be found in the group "Nature, Tools, Housing". The SMC list was constructed to demonstrate the phonological similarity within the Slavic family, disregarding the differences in meaning when they don't fit into the picture. The category, however, usually remains the same. This is the case with the word *gora*, but also of *cep*, the archaic word for 'chain' (elsewhere in Slavic 'flail'; in Bulgarian, replaced in this sense by *veriga*) or the adjective *zăl* 'angry' (elsewhere 'bad'; in Bulgarian, replaced in the general meaning by *loš*).

3. Another problematic issue is the alleged focus on the standard language in the SMC list. Meľnyčuk and Carlton weren't fully consistent in this aspect, as for example *člun*

---

2   In fact, the old word is rarely attested in dialectal *лес* or *ляш*, as well as in composite words like *лесничар* 'forester' (*Rečnik* III, p. 367).

'boat' and *pes* 'dog', reflexes for **člъnъ* and **pьsъ*, don't fully agree with the standard language's sound laws; rather, they reflect the phonological changes of peripheral dialects (c.f. Mladenov 1941, p. 419, 689), as far as we can consider these dialects (e.g. Torlak dialects) as parts of the Bulgarian linguistic area. Such words we consider as inherited, but they are rarely used in standard Bulgarian, which prefers *lodka* for 'boat' and *kuče* for 'dog'.

4. Moreover the list also doesn't consider possible borrowings between Slavic languages, which aren't observable based on phonological criteria, e.g. in the case of the word *pivo*: in Old Church Slavonic it is attested only with the general meaning 'drink', while in modern Bulgarian *pivo* means 'beer', perhaps under the influence of Czech[3]. A further curious case is the word *čužd* 'foreign, strange' (OCS *stuždъ* or *štuždъ* 'foreign'), which reflects Proto-Slavic **tjudju* (itself a probable borrowing) by sound laws of both Russian (**tj > č*) and Bulgarian (**dj > žd*) – a typical Church Slavonic word[4].

5. Finally, not even a standard language is immune to changes on the lexical level. The word *pivo*, although still used in brands and other specific contexts (e.g. composita like *pivovarna* 'brewery'), has mostly been replaced by the Italian loanword *bira* now.

For the aforementioned reasons, the SMC list requires certain modifications and improvements for our purposes. Sometimes it is unclear whether a word from the SMC list which is presented as a Bulgarian reflex of a Proto-Slavic lemma is actually related to it (e.g. the already mentioned word *dom*). Thus, in a first step we added the meanings to the particular lemmas in the Bulgarian column of the list, marking the cases in which the modern meaning was significantly different from that of an attested Old Church Slavonic (or an acceptable Proto-Slavic) cognate. In a further step we compared the meanings with the other Slavic languages as well. When meanings of Proto-Slavic and Old Church Slavonic lemmas differed from those of modern Bulgarian, we also added certain synonyms, based on descriptions of the lemmas provided by Bulgarian dictionaries (especially Mladenov 1941 and Georgiev/Račeva et al. 1971–2002).

As already mentioned, the original SMC list doesn't include Bulgarian reflexes for six of the Proto-Slavic (and Old Church Slavonic) lemmas: **bry*, **dęsna*, **lěsъ*, **medvědъ*, **mǫka* and **ǫgor'*. Although one could expect that they were replaced by a borrowing, this is not the case. The missing lemmas in the Bulgarian column

---

3   A similar situation is described for Slovene (cf. Kelih 2015, p. 31).

4   As a rule, words attested in Church Slavonic but lacking in dialectal material would be considered loanwords. However, in the case of *čužd* 'foreign, strange' the situation is unclear from this aspect, but as we are dealing here with an adoption of a foreign sound material (Russian or dialectal Serbian reflex of **tj > č*), the word can be considered as a borrowing.

of the SMC list are all either dialectally attested (*bărva*, *desna*, *ljaš*, *medved*[5]) or replaced by another inherited root (*vežda* 'brow' for *bry, *venci* 'gum' for *dęsna, *gora* 'forest' for *lěsъ, *brašno* 'flour' for *mǫka, *zmiorka*[6] 'eel' for *ǫgor'). For some of the meanings we included the synonyms common in dialects (e.g. *jagulja* for *zmiorka*) as well. Similarly, in the expanded list we have replaced the dialectal (e.g. *člun*) and unattested (e.g. *krasen*) forms with standard Bulgarian (*čăln*, *krasiv*) forms[7]. The modified list in the end includes a further 47 lemmas, added to the original grammatical and semantic categories:

(1) *hubav* 'beautiful', *običan* 'beloved', *goljam* 'big', *loš* 'bad',
(2) *petel* 'rooster', *kuče* 'dog', *gligan* 'boar', *dobităk* 'cattle', *zmiorka* (or *jagulja*) 'eel', *mečka* 'bear'
(3) *bakla* 'bean', *gora* 'forest', *kăpina* 'blackberry, raspberry bush', *hvojna* 'juniper',
(4) *tiča* 'run', *slušam* 'hear, listen', *jazdja* 'ride', *karam* 'ride' (a vehicle), *iskam* 'want', *vărvja* 'walk', *hodja* 'go', *trăgna* 'go (out)', *dviža se* 'move', *obeštavam* 'promise', *treperja* 'tremble', *gotvja* 'cook',
(5) *kum* 'godfather, best man', *badžanak* 'best man, brother-in-law', *hora* 'people', *săprug* 'husband', *bašta* 'father',
(6) *veriga* 'chain', *mlatilo* 'flail', *lodka* 'boat', *kăšta* 'house', *vrata* 'door', *karuca* 'cart', *ralo* 'plough'
(7) *kărmja* 'feed', *mălzja* 'milk', *bira* 'beer', *brašno* 'flour',
(8) *vežda* 'brow', *venci* (or *desna*) 'gum', *krak* 'leg'.

3 Analysis

The working hypothesis of our approach is the idea that the basic vocabulary (in our case the empirical data is the SMC list) is a stable lexical stratum, resistant

---

5 The standard Bulgarian word for 'bear' мечка is most likely a tabuised reflex of *medvědъ as well. Other dialectal forms are медведа or видмедиха (*Rečnik* III, p. 777).

6 Georgiev (cf. *Rečnik* I, p. 477) reports *egulja* and *jagulja* as the common words for 'eel' in western Bulgarian dialects. These are most probably early borrowings from a Romance language (Lat. *anguilla* 'eel'). The word *brašno* is old (OCS *brašъno* 'food'), but the proposed PIE root **bʰar-* is irregular for PIE, and also attested only in its western branches (e.g. Lat. *farīna* 'flour', OIc. *barr* 'grain', Wel. *bara* 'bread'), thus the word is often seen as an ancient borrowing from a European substrate language (c.f. Derksen 2008, p. 57).

7 For the full discussion on the modification of the original SMC list, cf. our forthcoming article on the topic (Kelih/Šimko 2018).

---

to borrowings[8]. Thus to challenge this hypothesis we searched for any potential loanwords in the given list. This requires a survey of available etymologies for the particular lemmas. The search for borrowings proceeds mostly by using a negative method: when the etymology points to an inherited root – in the ideal case a Proto-Indo-European one – the possibility of borrowing is disregarded. However, the etymological results aren't easy to quantify, because neither a borrowing nor inheritance are absolute categories. The status of a particular word as a "borrowed" or "inherited" one might be contested on both the synchronic and on the diachronic level.

We have already mentioned two phenomena which make it difficult to determine the status of a borrowing on the synchronic level, namely the integration of peripheral dialects into the standard language, and contact between multiple related languages. When we mark the words like *člun* or *pes* as "inherited", we implicate either an influence of foreign language on only a part of it, or a sound law affecting only a single lemma, thus violating the principles of sound change without exception. Yet they can't be classified as borrowings in the same way as the words like *konstitucija* 'constitution' or *hipermarket* 'hypermarket', which arguably didn't have any comparable cognates in the Bulgarian dialectal area before they were borrowed into the standard language.

The diachronic level opens even more questions. First of all, when does a word in fact become "inherited"? From an idiolectal point of view, most words, perhaps with the exception of childish utterances like *mama*, are borrowed. Many modern Bulgarian words are inherited from local dialects, like *bair* 'hill' or *hora* 'people', rather than from literary Church Slavonic, which has *gora* and *ljudje*; the dialects themselves have borrowed them. Thus we can state that they are inherited from pre-standard Bulgarian, but also that they are borrowings into pre-standard Bulgarian. It is also questionable whether calques built up from inherited roots like *mravojad* 'anteater' or *petiletka* 'five-year plan' can also be considered "inherited", as they most likely weren't present in earlier linguistic strata[9]. Linguistic inheritance is not a genetic relation, but rather a vector,

---

8 Carlton himself formulates the idea in the foreword to his *Phonological History* (1990, p. 6), where he points out the "remarkable similarity" of the basic vocabulary of Slavic languages.

9 Although these words weren't part of the SMC list, 'the anteater' can in fact be found among the meanings in WOLD. A similar case could be *nevesta* 'bride' (OCS *nevěsta*): the Slavic negative prefix is attached to a reflex of the PIE root **uoid-t-* (c.f. Derksen 2008, p. 351), so it would literally be 'the unknown one'. The construction isn't found outside Slavic languages, and thus we can surely say it is inherited only from Proto-Slavic. We may reconstruct a PIE form, but we cannot say with certainty

pointing at a reference point in the past. We can speak of words inherited from early Bulgarian, Proto-Slavic or Proto-Indo-European, thus disregarding the possibility of an earlier borrowing.

The analysis thus focused on the following points. First, we looked for the earliest reconstructible Bulgarian (or Slavic) form: this established the basic answer to the question of whether it is inherited or borrowed. Second, we have replaced this dichotomy of borrowing/non-borrowing with a scale similar to the one used in WOLD (Haspelmath/Tadmor 2009b)[10]: from "1" (clearly borrowed) to "5" (no evidence for borrowing):

1 clearly borrowed
2 likely borrowing with a known donor, plausible etymology as a borrowing
3 multiple arguments for borrowing, but the donor is unknown
4 most likely an inherited word, irregularities in reconstruction of protoform
5 no evidence of borrowing.

In the original SMC list, we identified 23 possible loanwords:

Lemma	Status	Stratum[a]	Donor
buk 'beech'	1	Late pre- or post-PSl	Germanic
vino 'wine'	1	Late pre- or post-PSl	Germanic or Romance
plug 'plough'	1	Post-PSl	Germanic
hljab 'bread'	1	Late pre- or post-PSl	Gothic
čereša 'cherry'	1	Late pre-PSl	Germanic
čužd 'foreign, strange'	1	Modern BG	Church Slavonic
kopăr 'dill'	2	Late pre- or post-PSl	Romance
skot 'cattle'	2	Late pre- or post-PSl	Gothic
*tjudju 'strange'	2	Late pre- or post-PSl	Germanic
kobila 'mare'	3	Early pre-PSl	unknown

this composition is that old, unlike, for example, jastreb 'hawk', which is likely cognate to lat. accipiter 'hawk, falcon', both likely reflecting PIE *h₁oh₁k̑u-ptr-, literally 'fast flier' (cf. Derksen 2008, p. 29). However, it also doesn't dismiss the possibility that the word is a calque based on a foreign word. In WOLD (cf. Haspelmath/Tadmor 2009, p. 14) calques, borrowings of mere semantic material, were originally not marked at all, because they were created in recipient languages.

10 The numbering reflects the one employed by Haspelmath and Tadmor in WOLD. The original proposal for the database project (cf. Haspelmath/Tadmor 2009a, p. 13) uses a scale from 0 (no evidence for borrowing) to 4 (certainly borrowed).

Lemma	Status	Stratum[a]	Donor
koza 'goat'	3	Pre-PSl	Turkic (?)
tur 'aurochs'	3	Pre-PIE or early pre-PSl	Semitic (?)
brada 'beard'	4	Pre-PSl	Germanic (?)
brjag 'coast, shore'	4	Pre-PSl	unknown
elha 'fir'	4	Early pre-PSl	unknown
esetra 'sturgeon'	4	Pre-PSl	unknown
gabăr 'hornbeam'	4	Early pre-PSl	unknown
găska 'goose'	4	Pre-PSl	unknown
kon 'horse'	4	Pre-PSl	unknown
krava 'cow'	4	Early pre-PSl	Celtic (?)
kruša 'pear'	4	Early pre-PSl	unknown
mljako 'milk'	4	Pre-PSl	unknown
smreka 'spruce'	4	Pre-PIE or early pre-PSl	unknown
jabălka 'apple'	4	Pre-PIE or early pre-PSl	unknown

[a]The borrowings are historically situated relatively, by periods marked by characteristic sound changes, attestations and historical context. The newest stratum of the vocabulary is "modern Bulgarian" (BG), containing the words unattested in Middle Bulgarian (until ca. 16th century). The earlier strata are defined mostly by more or less arbitrarily chosen sound laws, common to the whole linguistic clade; "early Bulgarian" is separated from post-Proto-Slavic (PSl) by the assibilation (*tj, *dj > št, žd), common to all Bulgarian dialects; post- from pre-Proto-Slavic by the first palatalisation (*k, *g, *x before a front vowel > č, ž, š), last major sound change common to all Slavic languages; and early pre-Proto-Slavic from the later period by Winter's law (emergence of an acute or laryngeal before a PIE media consonant), which has affected the Baltic languages as well. Under "Proto-Indo-European" we mean the stage of the language before the loss of difference between the laryngeals.

The modified list included 13 further possible borrowings:

Lemma	Status	Stratum	Donor
bakla 'bean'	1	Modern BG	Ottoman Turkish
badžanak 'brother-in-law'	1	Modern BG	Ottoman Turkish
bira 'beer'	1	Modern BG	Italian
karam 'drive'	1	Early BG	Romance
karuca 'cart'	1	Modern BG	Greek or Romanian
săprug 'husband'	1	Modern BG	Church Slavonic

Lemma	Status	Stratum	Donor
hora 'people'	1	Early BG	Greek
hubav 'beautiful'	1	Modern BG	Ottoman Turkish
jagulja 'eel'	1	Early BG	Romance
kum 'godfather, best man'	2	Post-PSl	Romance or Turkic
kuče 'dog'	2	Post-PSl	Turkic
gotvja 'cook'	3	Late pre- or post-PSl	Gothic (?)
brašno 'flour'	4	Early pre-PSl	unknown

The words for which we could easily reconstruct Proto-Indo-European (PIE) roots were marked with a 5. We also gave a 5 in those cases where only a common Balto-Slavic (e.g. *răka* 'hand', cf. Derksen 2008, p. 439) or Slavic (e.g. *riba* 'fish', cf. *Rečnik* VI, p. 245) root can be reconstructed, with unknown cognates in other related languages, so far as they don't show any irregularities from the aspect of morphological and phonetic developments. Unlike WOLD (cf. Haspelmath/ Tadmor 2009a, p. 13), we didn't mark the probable borrowings from substratum languages into PIE (e.g. *jabălka* 'apple' or *tur* 'aurochs') with a 5 if we couldn't determine whether the word was borrowed into PIE or into later strata.

The number and plausibility of arguments were decisive for the further status marks. This affects most of the pre-Proto-Slavic borrowings. The word *elha* 'fir', when compared with ahd. *elira* or lat. *alnus*, points at a root $*alis-eh_2$ or $*als-eh_2$ (cf. de Vaan 2004, p. 34, Derksen 2008, p. 370). Such a variation is untypical for Proto-Indo-European, and thus the lemma receives a 4. If more arguments for a borrowing – or against the inheritance from PIE – were present, the word was marked with a 3. The word *kobila* 'mare' also seems to be a part of a deeper Proto-Slavic stratum of vocabulary. There are more arguments for its status as a borrowing than in the case of *elha* – a comparison with lat. *caballus* points to the presence of (for PIE phonetics) a controversial $*a$; the second consonant points to a $*b$, which should have fed Winter's law (cf. Derksen 2008, p. 232). Furthermore, the suffix $*-yla$ is very rare in Slavic, elsewhere found only in the substrate loanword *mogila* 'burial mound' (*Rečnik* II, p. 501). The verb *gotvja* 'to cook, prepare', added to the extended list as a replacement of *varja* (which means only 'to boil' in contemporary Bulgarian), is often considered an early borrowing from Gothic *gataujan* 'to make', but it could also be a native reflex of the Proto-Indo-European root $*g^weh_2$- (cf. Pronk-Tiethoff 2013, p. 192). Both explanations are characterised by irregular developments, thus resulting in a status of 3 for the lemma.

The higher marks were given to those words where the arguments for borrowing are more robust. In the case of *skot*, an archaic word for 'cattle' in the original SMC list, we can also speak of a borrowing from Gothic *skatts* 'money'. Another etymology sees the Germanic word vice-versa as a borrowing from Slavic, where the root should reflect Proto-Indo-European $*skop-t-$ 'castrated' (*Rečnik* VI, p. 787). We mark this lemma with a 2, because the explanation of the lemma as inherited in Slavic encounters more obstacles than the former one (cf. Pronk-Tiethoff 2013, p. 144). Finally, the words where no plausible etymology as inherited roots could be offered, e.g. *čereša* 'cherry' (cf. Vasmer 1964 IV, p. 343) or *vino* 'wine' (*Rečnik* I, p. 149), were marked as "clear borrowings". In the final results, only lemmas marked as clear (1) or likely (2) borrowings were taken into consideration.

4 Results

The etymological survey shows clearly that the determination of the status of a lemma as borrowed often requires an analysis of the Proto-Slavic or even earlier form of the root. This problem, however, shouldn't lead us astray from the very fact that the basic vocabulary does include some loanwords. This by no means disproves the idea of the stability of this stratum. The old loanwords may indicate a historical situation of intense language contact (Gołąb 1992), but that also provides us with data about their longevity. Finally, the survey showed that the basic vocabulary is altered mostly by synonyms which penetrate and replace the inherited roots – words like *plug* or *hljab* in the original SMC list[11], or modern Bulgarian words like *bira*, *karuca* or *hora* in the modified version. Thus, the general result is that the basic vocabulary indeed incorporates selected foreign words, however it is a conservative, rather than an impervious lexical stratum.

Now we can turn to some of the details of our study. As previously mentioned, the original SMC list lacks six Bulgarian reflexes for Proto-Slavic roots ($*bry$,

---

11 The meaning of both words was most likely different in donor languages, receiving a general meaning after the borrowing. Germanic source of *plug* (< $*plōg$-) seems to have denoted a 'heavy plough', which was in use in Central Europe before the arrival of the Slavs. After the word was adopted into Slavic, such ploughs gradually replaced the hand ploughs or "ards" (e.g. PSl $*ar^{?}dla$ > Bulg. *ralo*) both in actual agriculture and in the terminology, receiving the general meaning 'ploughing instrument' (cf. Pronk-Tiethoff 2013, p. 93). Similarly, the Gothic *hlaifs*, the likely source of the word *hljab*, meant 'slice', similar to Slavic $*kruxъ$ (< PSl $*krauxu$). The semantic shift 'slice' > 'bread' appeared in both roots (S.Cr. *krŭh* 'bread').

*dęsna, *lěsъ, *medvědъ, *mǫka, *ǫgor'). However, there are attested reflexes of these roots in contemporary or older Bulgarian dialects as well. Thus we could identify with certainty only one borrowing in modern Bulgarian, which is the Church-Slavonicism *čužd* 'strange, foreign'. Most of the certain loanwords in the list were borrowed before the development of separate Slavic languages – *buk* 'beech', *vino* 'wine', *kopăr* 'dill', *plug* 'plough', *skot* 'cattle', *hljab* 'bread', *čereša* 'cherry', and most likely *tjudju* 'strange, foreign', the Slavic protoform of *čužd*, as well. Thus we can speak of eight likely loanwords only, or 3.77 % of the given core vocabulary[12]. Six of these seem to be borrowed from older Germanic languages.

The proposed modification of the list complemented missing reflexes and archaisms by synonyms, which can be found in standard Bulgarian and its major dialects. The resulting list of 253 lemmas included a further nine very likely loanwords – *bakla* 'bean', *badžanak* 'brother-in-law, best man', *bira* 'beer', *karam* 'ride, drive', *karuca* 'cart', *kuče* 'dog', *hora* 'people', *hubav* 'beautiful' and *jagulja* 'eel', complementing in their respective meanings the words *bob*, *dever*, *pivo*, *jazdja*, *voz*, *pes*, *ljude*, *krasen* and the missing reflex for *ǫgorъ. Furthermore, for the secondary meaning of *dever* 'best man' we have also added the synonymous lemma *kum* 'godfather, best man'; a similar addition is *săprug* 'husband' for *măž*. From the rest of the words, *bob* and *krasiv* are used nearly synonymously alongside *bakla* and *hubav*[13]. The words *jazdja* and *karam* have different meanings: *karam* is used only with inanimate forms of transportation like carts and cars, while *jazdja* canonically refers to transportation with animals such as horses and donkeys. Finally, the words *bira*, *karuca*, *kuče* and *hora* have replaced their former semantic equivalents in their general meaning. Thus we get in total 19 loanwords, or 7.5 % of the core vocabulary. Most of the newly added loanwords are specific for Bulgarian; only *kum* and *kuče* are attested in other Slavic languages as well.

The status of borrowings doesn't have to reflect the depth of stratum. The original SMC list included eight loanwords, which (excluding *čužd*) can all be found in other Slavic languages too. The modified list shows another two likely loanwords in this stratum, replacing inherited roots in Bulgarian. One word (*čereša*) was surely borrowed even into Pre-Proto-Slavic, as it has undergone the first palatalisation. The numbers before and after the separation of Bulgarian roughly correspond. Thus the linguistic contact between earlier Slavic and its Germanic and Turkic donors wasn't very different from the later Bulgarian contact with Ottoman Turkish, Greek and Romance languages.

As already mentioned, the results exclude the less certain borrowings, comprising 15 words in the original SMC list (*brada* 'beard', *brjag* 'coast, shore', *elha* 'fir', *esetra* 'sturgeon', *gabăr* 'hornbeam', *găska* 'goose', *kobila* 'mare', *kon* 'horse', *koza* 'goat', *krava* 'cow', *kruša* 'pear', *mljako* 'milk', *smreka* 'spruce', *tur* 'aurochs' and *jabălka* 'apple') and a further two from the revised list (*brašno* 'flour' and *gotvja* 'to cook, prepare'). The status of these words as "uncertain borrowings" is in most of these cases an indication of a problematic or irregular reconstruction of their Proto-Indo-European roots (or, vice-versa, of the borrowed words) from a phonological, accentological or morphological point of view. Only *gotvja* has an identified donor language (Gothic). In other cases we can rarely define even the language family.

5 Conclusions

Every analysis of this type shows us some aspects of the local etymological tradition. In some cases it is more a political than linguistic question where "Bulgarian" ends and "Macedonian", "Serbian" or "Church Slavonic" begins. Some words in the original SMC list (e.g. *člun*, *pes*) show us that Modern Bulgarian isn't based on a single dialect. Along with many other standard languages in general, it is an integrative construct which attempts to include a broad field of dialects (especially in the case of words for animals and plants) and sociolects (especially Church Slavonic and administrative Russian – see the words like *čužd* and *săprug*). It is surely not a rigid, closed language, and this fact is reflected in the observed number of loanwords in its core vocabulary.

On the other hand, the study also opened multiple perspectives for our project concerning the topic of South Slavic loanwords. First, it offers a method for the determination of probable borrowings, which has been reflected in theoretical works of the WOLD project (Haspelmath/ Tadmor 2009b), but the studies themselves lacked precise criteria for determining the particular loanword status. Second, it opens the question of semantic drift of lemmas, which helps us to clarify the supposed context of contact situation and also the loanword status itself. Third, it adopts a framework for a closer analysis of prehistorical contact situations by researching the earlier, less certain borrowings, as well as modern-era dialects. The project thus harnessed itself for the study of a larger sample – and for bringing South Slavic etymology to the digital era.

---

12   The earlier studies of Slovene (Kelih 2015) and Croatian (Kelih/Garić 2016) show us only slightly different results, namely 13 for Slovene and 14 for Croatian. This quantitative difference rather reflects the readiness of the author to accept the less clear borrowing status than some in-depth substantial differences between these South Slavic languages.

13   Carlton mentions *krasen*, which has of course the same root, but the suffix isn't used today in the standard language.

References

Carlton, Terence: *Introduction to the phonological history of the Slavic Languages*. Columbus: Slavica. 1990.

Cejtlin, Ralja (ed.): *Staroslavjanskij slovar'*. Moskva: Russkij jazyk. 1994.

D'Andrade, Roy: *The development of cognitive anthropology*. Cambridge: Cambridge University Press. 1995.

Derksen, Rick: *Etymological Dictionary of the Slavic Inherited Lexicon*. Leiden/Boston: Brill. 2008.

Embleton, Sheila M: *Statistics in historical linguistics (Quantitative Linguistics 30)*. Bochum: Brockmeyer. 1986.

Georgiev, Vladimir/Račeva, Maria (eds.): *Bălgarski etimologičen rečnik*. Sofia: BAN: 1971 (t. I), 1979 (t. II) 1986 (t. III), Sofia: Marin Drinov: 1995 (t. IV), 1999 (t. V), 2002 (t. VI).

Gołąb, Zbigniew: *The Origins of the Slavs: A linguist's view*. Bloomington: Slavica. 1992.

Haarmann, Harald: "Basic" Vocabulary and Language Contacts: The Disillusion of Glottochronology. In: *Indogermanische Forschungen* 95. 1990, pp. 1–37.

Haspelmath, Martin/Tadmor, Uri: *Loanwords in the world's languages: A comparative handbook*. Berlin: de Gruyter. 2009a.

Haspelmath, Martin/Tadmor, Uri (eds.). *World Loanword Database*. Leipzig: Max Planck Institute for Evolutionary Anthropology. 2009b. Retrieved 11.06.2017 from http://wold.clld.org.

Hock, Hans Henrich/Joseph, Brian D.: *Language history, language change, and language relationship. An introduction to historical and comparative linguistics (Trends in Linguistics. Studies and Monographs 93)*. Berlin, New York: Mouton De Gruyter. 1996.

Ilčev, Stefan: *Rečnik na redki, ostareli i dialektni dumi v literaturata ni ot XIX i XX vek*. Sofia: IK "Emas". 1998.

Kelih, Emmerich: Problemen der empirischen Lehnwortforschung: Lehnwörter im Basiswortschatz (Slovenisch) und ein frequenzbasierter Ansatz (Kroatisch). In: Kelih, Emmeric/Fuchsbauer, Jürgen/Newerkla, Stefan M. (eds.): *Lehnwörter im Slavischen: Empirische und crosslinguistische Perspektiven*. Wien: Universität Wien. 2015.

Kelih, Emmerich/Krebs-Garić, Catharina: Lehnwörter im kroatischen Basiswortschatz (Swadesh-Liste) und ein Vergleich Kroatisch – Slowenisch: Eine Pilotstudie. *Anzeiger für Slavische Philologie* 44. 2016, pp. 9–25.

Kelih, Emmerich/Šimko, Ivan: *Lehnwörter im bulgarischen Basiswortschatz (Swadesh-Liste) – eine Pilotstudie* (in print). 2018.

Mladenov, Stefan: *Etimologičeski i pravopisenъ rečnikъ na bъlgarskija knižovenъ ezikъ*. Sofia: Ch. G. Danov. 1941.

Pronk-Tiethoff, Saskia: *The Germanic Loanwords in Proto-Slavic*. Amsterdam/New York: Rodopi. 2013.

Swadesh, Morris: Lexico-Statistic Dating of Prehistoric Ethnic Contacts. With Special Refernce to North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96 (4). 1952, pp. 452–463.

Tadmor, Uri: Loandwords in the world's languages: Findings and results. In: Haspelmath, Martin/Tadmor, Uri (eds.): *Loanwords in the world's languages. A comparative handbook*. Berlin et al.: de Gruyter. 2009, pp. 55–75.

Tadmor, Uri/Haspelmath, Martin/Taylor, Bradley: Borrowability and the notion of basic vocabulary. *Diachronica* 27 (2), 2010, pp. 226–246.

Vasmer, Max: *Etimologičeskij slovar' russkogo jazyka*. Transl. O.N. Trubačev. Moskva: Progress. 1964.

Zenner, Eline/Speelman, Dirk/Geeraerts, Dirk: Core vocabulary, borrowability and entrenchment: A usage-based onomasiological approach. *Diachronica* 31 (1), 2014, pp. 74–105.