LINGUISTIK
INTERNATIONAL

Luka Szucsich / Agnes Kim /
Uliana Yazhinova (eds.)

Areal Convergence in Eastern
Central European Languages
and Beyond

PETER LANG

This book assembles contributions dealing with language contact and areal linguistics. The goal of the book is to investigate linguistic convergence in Europe with a strong focus on the languages of Eastern Central Europe which show many remarkable similarities. The focus is put on a methodical and empirical component in the investigation of two or more languages in the context of possible language contact phenomena. Languages of Eastern Central Europe and adjacent parts of Europe use a considerable amount of common vocabulary due to the transfer of loanwords during a long period of cultural contact. But they also share several grammatical features—phonological, morphological and syntactic ones. This book tackles lexical and grammatical phenomena in language contact situations. The authors take up diachronic, synchronic and language acquisitional perspectives, and discuss methodological problems for the field.

Luka Szucsich is a professor of East Slavic linguistics at the Department of Slavic and Hungarian Studies of the Humboldt University of Berlin.
Agnes Kim works as a research assistant in the Special Research Programme (SFB) "German in Austria" at the Department of Slavonic Studies of the University of Vienna.
Uliana Yazhinova is a researcher assistant for East Slavic linguistics at the Department of Slavic and Hungarian studies of the Humboldt University of Berlin.

Areal Convergence in Eastern Central European Languages and Beyond

LINGUISTIK
INTERNATIONAL

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

BAND 44

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

AREAL CONVERGENCE IN EASTERN CENTRAL EUROPEAN LANGUAGES AND BEYOND

PETER LANG

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

Luka Szucsich, Agnes Kim, Uliana Yazhinova

Introduction

This volume assembles written versions of contributions presented at a workshop organized at the Slavic Department of the Humboldt University Berlin. The workshop was conducted within the project *Areal Convergence in Eastern Central European Languages* (ACECEL), which was part of the CENTRAL network (Central European Network for Teaching and Research in Academic Liaison). CENTRAL was initiated by the Humboldt University Berlin and founded together with the Universities of Vienna and Warsaw, Charles University Prague as well as ELTE Budapest with the goal to establish joint projects for the promotion of exchange in research and teaching. All CENTRAL projects were funded by the German Academic Exchange Service (Deutscher Akademischer Austauschdienst, DAAD) for four years (2015–2018).[1] The primary goal of ACECEL was to investigate linguistic convergence in the languages of Eastern Central Europe which show many remarkable similarities. Luka Szucsich (Humboldt University Berlin) and Stefan Newerkla (University of Vienna) were the PIs of ACECEL which—at different stages—also included the involvement of colleagues in Prague, Warsaw and Budapest. The focus was put on a methodical and empirical component in the investigation of two or more languages in the context of possible language contact phenomena. The partner institutions with focusses in the field of lexical borrowing, in particular for German, Slovak and Czech (Slavic Department, University of Vienna) and in the field of morphosyntactic typology and microvariation in the field of Slavic languages (Slavic Department, Humboldt University) provided an ideal basis for cooperation for the described endeavors. The workshop—although still mainly focusing on Eastern Central Europe—took a broader areal perspective including larger parts of Europe, which is also reflected in some of the volume's contributions (cf. Stolz/Levkovych, Šimko/Kelih and Tetereva/Naukhatskaia).

Languages of Eastern Central Europe and adjacent parts of Europe use a considerable amount of common vocabulary, which is certainly due to the transfer of loanwords during a long period of cultural contact (cf. the contributions of Newerkla, Šimko/Kelih and Tölgyesi). But they also share several interesting

---

1   For more information on the network and its programs see https://www.projekte. hu-berlin.de/en/central/network/index_html.

and distinct grammatical features ranging from (i) valency and government patterns (cf. Gaszewski, Kim and Kim/Scharf/Šimko), (ii) modality systems (cf. Martínek) and (iii) morphosyntactic patterns (e.g. doubly-filled-comp properties, prefixal verbal composition, purposive infinitival clauses, cf. Szucsich), to name just a few structural properties. Furthermore, Stolz/Levkovych investigate the distribution of BELONG-constructions in a broader European context. Last but not least, Tetereva/Naukhatskaia discuss the acquisition of verbs and verbal categories in Russian-German bilinguals.

Some of the contributions take up a diachronic (cf. Kim and Martínek) and/or a theoretical perspective (cf. Januška, Kim and Newerkla), discussing methodological as well as metalinguistic issues. One of the problems which is discussed in a lot of the papers of this volume, and which still remains unresolved is a clear understanding of the processes involved in language contact and how they may form what has been labelled a linguistic area.

Two of the papers published in this book (cf. Tetereva/Naukhatskaia and Kim/Scharf/Šimko) present results of two subsequent international cooperation projects funded by the CENTRAL network in 2016 and 2017, namely the CENTRAL-Kollegs *Empirical perspectives on area-typological aspects of language contact and language change* and *From language contrast to language contact: Corpus linguistic approaches to language contact phenomena*. This format for research based-learning aimed at junior researchers and students. Under the guidance of junior researchers, student research teams developed and carried out short-time research projects. The two above mentioned CENTRAL-Kollegs were both organised by Uliana Yazhinova (Humboldt University Berlin), Agnes Kim (University of Vienna) and Karolína Vyskočilová (Charles University Prague). Eight BA- or MA-students participated in each of them; two from Vienna, two from Prague and four from Berlin.

The main idea of the two abovementioned projects was (i) to create a framework within which the participating students could conduct their own research, and (ii) to bridge the gap between studying and research in Slavic linguistics. In order to facilitate these learning and research processes the organisation team equally pursued and connected didactic and scientific goals in both CENTRAL-Kollegs. The program was based on the so-called *research competence model*, which comprises two main components: (i) language contact and convergence as scientific objectives, and (ii) guided research and mentoring as teaching strategies. The main target was to guide the students throughout the whole research process as depicted in Fig. 1 and thus support them in developing linguistic and social competences to plan and conduct collaborative research projects and maybe even pursue a scientific career.
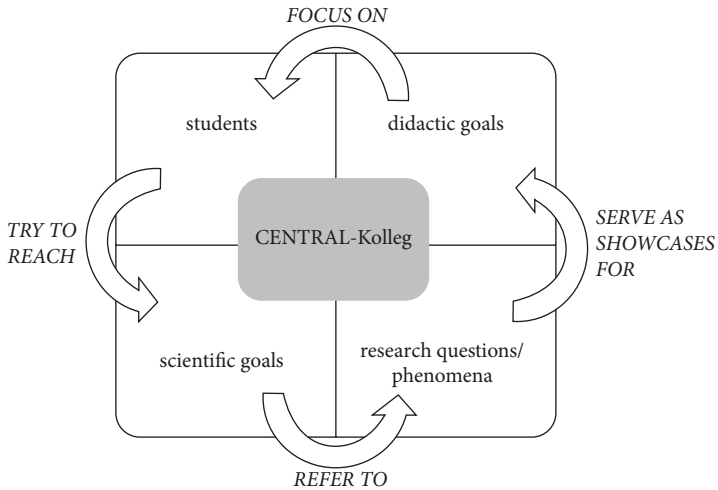
Fig. 1: Guided research in the CENTRAL-Kollegs

The students collaboratively investigated selected language contact and convergence phenomena by employing recent corpus linguistic methods and thereby added to the research of the CENTRAL project ACECEL. Additionally, the projects as a whole contributed to the dissemination of up-to-date empirical methods in Slavic contact linguistics. Thus, we hope that this book represents a successful example of a collaborative endeavor brought about by senior researchers, junior researchers and students.

Jerzy Gaszewski

Does Verb Valency Pattern Areally in Central Europe? A First Look

Abstract: The paper reports on the results of a pilot study of verb valency in Central Europe. The analysis relies on semantic roles associated with individual predicates, so-called *microroles* (Hartmann/Haspelmath/Cysouw 2014), and involves a comparison of the distribution of grammatical elements (cases and adpositions) marking the arguments corresponding to particular microroles. The degree to which the distributions of markers in certain languages correspond to each other is captured by means of a distance index, a simple statistical measure introduced in the paper. The results provide some support for *Donaubund* (represented by Czech, German and Hungarian), but also show general parallels among all the investigated languages.

Keywords: verb valency, argument marking, language distance measurement, linguistic areas, Danube Sprachbund

1 Introduction

Central European languages have been researched from an areal perspective for a considerable time. The field of areal linguistics itself is currently shifting away from determining whether "a given geographical area could be classified as a linguistic area or not" (Hickey 2017, p. 2) to describing particular areal patterns of language structure as such (cf. also Campbell 2017, and, with particular reference to Central Europe, Januška this volume). In this vein, I seek to investigate how one grammatical phenomenon, verb valency, patterns in the region.

The main objective of this paper is to give a preliminary answer to the question posed in the title, based on data from a pilot study.[1] Apart from that, the

---

paper aims to lay down the conceptual foundations which further research will inherit from the pilot study, with various improvements (cf. section 4.3). Thus, another objective of the paper is to present the project's theoretical approach to the comparison of valency across languages (section 2.3) as well as the methods of data analysis (section 3.2).

The present paper deals with a limited group of five languages, chosen for the pilot study: Czech, German, Hungarian, Polish and Romanian. It is clear that the group must be broadened in later research (cf. section 4.3). The choice of languages was partly dictated by the availability of informants, but I also intended to obtain a graded sample of Central Europe.

The region has been variously delineated by linguists (overviews can be found, e.g., in Kurzová 1996, Newerkla 2000, Newerkla 2002, Pusztay 2015 and Januška this volume). The region has been referred to by different names as well. *Central Europe* is the most commonly used label, but we also encounter *Donaubund*, *Carpathian Sprachbund* (Thomas 2008 after Januška this volume), *Mitteleuropa* and *Zentraleuropa* in German-language sources, or *Amber Road Region* (Pusztay 2015). Usually a separate label is combined with a proposal concerning alteration of the membership or structure of the group. The different concepts will not be discussed at length here, the reader is referred to the works mentioned above. Among the many proposals, the most widely accepted seems to be the idea of a linguistic area encompassing Czech, Slovak, Hungarian and (Austrian) German as the core languages, with Polish, Slovenian and Croatian as marginal members. I will use the label *Donaubund* to refer to this grouping and *Central Europe* will be reserved for the broader region ranging from the Baltic Sea to the Balkans and from German-speaking countries to Russia.

The five selected languages represent different levels of established "Central Europeanness". Czech and German are very often grouped together by areal studies and can be said to belong to the core of the region. Of the two, German has generally been a major superstrate throughout the region. Hungarian is also typically grouped with the former two languages, although it is genetically unrelated to them (or any of its neighbours).[2] Any observed similarities

---

2   In our data the otherness of Hungarian is visible in the forms of valency markers—it has by far the largest repertoire of case suffixes and, as the only language in the group analysed here, has postpositions. These features are a Finno-Ugric legacy and have nothing to do with areal similarity in Central Europe. In any case, I am interested in the distribution of the markers rather than their form.

involving Hungarian make a good case for areal convergence. The position of Polish is much more ambiguous. It is not always included in studies on Central Europe and it is never considered to belong to the core. Somewhat confusingly, it shares similarities on all levels of language structure and in the very form of linguistic units with Czech, which is due to a close genetic relationship. Common origin, the "default" reason for Czech-Polish similarities is a confounding factor for an areal study. I propose a way of controlling for this in the analysis (cf. section 3.4). Romanian, the last of the five languages, is geographically proximate, but typically grouped within the Balkan linguistic area and not Central Europe. It is included as an expected natural outlier. Apparently, Romanian is only ever considered together with Central European languages when Balkan languages in general are included in the given study. This is evident when one traces the few mentions of Romanian throughout the overview text by Januška (this volume).

Using such a graded sample of Central Europe, the extent of the region is consciously left underspecified. Ultimately, it is the data that will show how the languages of the region relate as far as verb valency is concerned. Consequently, the above description of the five languages and their status within Central Europe is only a point of reference, a picture we get from earlier research on the area, much of which does not even deal with valency.

Any language possesses hundreds if not thousands of valency structures. It is impossible to investigate all of them and, hence, selection of material is necessary. The choices made in the course of this selection are of great importance. However, there are no generally approved methods to obtain a sample of verbs of a language for a comparative study (cf. Haspelmath 2015, p. 134). The problem is addressed in full in Gaszewski (in preparation b). Here, it is sufficient to state that several broad groups of verbs are dispreferred in the study. The reasons for that are explained in section 2.2. The 75 verbal meanings selected for the pilot study are provided in Appendix 1.

The data was gathered in the form of example sentences from five people, each a native speaker of a different language. Informants are obviously only one of the possible sources of data on valency. The advantages and disadvantages of each type of source are discussed extensively in Gaszewski (in preparation a), the ultimate choice being to have informants as the main source of data throughout the project, with help from other types of sources whenever possible. In the pilot study the main benefit of informants was that they allowed the gathering of a substantial amount of data relatively quickly and easily.

2 Valency and the Comparison of Valency

2.1 The Notion of Valency

Valency refers to the property of some lexical items (*valency carriers*) which require other items to co-occur with them to complete their meaning.[3] In other words, valency carriers open slots (*valency positions*) to be filled by dependent words. Let us now analyse how this works in a single sentence.

(1) Kupiłem   od   sąsiada   samochód za milion. (Pl, 2)[4]
     buy.PST.M.1SG from neighbour.GEN car.ACC   for million.ACC
     'I bought a car from my neighbour for a million.'

The central element in the clause, the valency carrier, is the verb *kupiłem* (an inflected form of *kupić* 'to buy, PFV'). It opens several valency positions filled by the phrases in the sentence, i.e., the arguments of the verb.[5] They are: *samochód* 'a car' (the product bought), *za milion* 'for a million' (the price paid), and *od sąsiada* 'from (a/the) neighbour' (the seller). The last valency position is the buyer, the subject of the clause, expressed in (1) solely by the 1SG marking on the verb. An explicit subject is possible, but not necessary in Polish.

The whole structure consisting of the valency carrier and the dependent valency positions is called a *valency frame*.[6] Note that the semantic roles above are defined with respect to a single predicate 'to buy'. In accordance with the

---

3 This formulation of the definition of valency is based directly on Karolak's (1999, p. 629) encyclopedic entry, but ultimately it stems of course from Tesnière (1959). The general idea that some words open slots to be filled by other words arose even earlier (cf. Bühler 1934).

4 Throughout the paper I employ the following abbreviations for the languages: 'Cz' for Czech, 'G' for German, 'H' for Hungarian, 'Pl' for Polish, 'R' for Romanian. Examples taken from the pilot study data also have numbers referring to sentences used for data collection, which are listed in Appendix 1.

5 Note that the idea of the *completion of meaning* of a valency carrier does not entail obligatory realisation of all valency positions in all real language occurrences of the valency carrier. This has already been noted by Tesnière (cf. the discussion in Ágel/ Fischer 2009, p. 230). On the contrary, verb valency is only one of the factors that influence the actual make-up of clauses (cf. Herbst/Heath/Roe/Götz 2004 on different kinds of "necessity" of phrases within clauses). Indeed, some valency positions are typically optional arguments, e.g., in (1), the arguments of price and seller.

6 Tesnière's (1959) original approach associated valency exclusively with verbs. Linguists have since realised that not only verbs can be valency carriers. Still, valency frames of verbs as templates of clauses will be the main focus of this paper.

terminology of Hartmann/Haspelmath/Cysouw (2014), I will call roles at this level of specificity *microroles*.

Each argument realises a semantic role related to the meaning of the valency carrier. However, we are able to identify the arguments and interpret the clause correctly thanks to the morphosyntactic marking of the individual phrases, covered mainly by cases and adpositions in (1). These kinds of grammatical markers are especially prominent as means of marking arguments in the languages under scrutiny. They will, therefore, be main objects of analysis here. Naturally, valency may be marked in other ways too (e.g., head-marking, element order).

2.2 Valency in the Areal Perspective

It is worth asking to what extent valency is conducive to areal influence. Before we refer to sources considering the matter with regard to language structure in general, let us briefly focus on Central Europe. Some studies (e.g. Newerkla 2000, p. 11, Pusztay 1996 after Pilarský 2001, Bláha 2015, p. 156–157) see the parallel use of valency markers as an areal feature of Central Europe. Probable calques in the use of prepositions in non-standard varieties of German had been noted throughout the 19th century and collected by Schuchardt (1884; after Kim this volume). Such observations are promising from our perspective. By contrast, Pilarský (2001: 118) staunchly rejects the validity of parallels in valency marking among Central European languages. Yet, most of the contemporary studies quoted above (except for Bláha 2015) consider a very limited number of valency structures, and so a broader analysis appears necessary.

It should also be made clear that the phenomenon of valency is rich (cf. section 1) and not uniform. Many valency structures show high regularity within (and across) languages. This regular side of valency is transitivity. In other structures, the marking of valency positions is word-specific (lexically governed by individual valency carriers). The question underlying the discussion below is which aspects of valency are of primary interest for an areal study. The answer has implications for the choice of verbs to be included.

Tab. 1 shows two rankings of levels and elements of language structure according to the likelihood of being affected by areal patterns. Both rankings are very general and neither mentions valency explicitly. However, we can relate some points in the rankings to different types of valency structures. The marking associated with individual verbs is essentially idiomatic and we can link it with the "structure of idioms" at the top of Aikhenvald's (2006) scale. Hickey (2017) puts "vocabulary" in the top position, but this point includes "phrases", which allows us to link the same kind of valency structures with it.

Tab. 1: Rankings of elements and levels of language according to conduciveness to areal patterning

Aikhenvald (2006, p. 5)	Hickey (2017, p. 6)
more similar to neighbouring languages	*levels most affected* [by areal influence (JG)]
Structure of idioms	Vocabulary (loanwords, phrases)
Discourse structure	Sounds (present in loanwords)
Syntactic construction types	Speech habits (general pronunciation, suprasegmentals)
Core lexicon	Sentence structure, word-order
Inflectional morphology	Grammar
more similar to genetic relatives	*levels least affected* [by areal influence (JG)]

Note: The original vertical arrangement of Aikhenvald's ranking is reversed here.

Transitivity can be associated with items that are lower in the rankings: "syntactic construction types" (middle of scale in Aikhenvald's ranking) and "sentence structure" (below the middle point on Hickey's scale). The regular side of valency is thus less likely to show areal influence.

This affects a significant number of verbs. Both plain intransitives (monovalent verbs with the marking of the sole argument predictable from the language's general alignment type) and plain transitives (bivalent verbs with similarly regular marking of arguments) appear to be of little interest for our study. Another reason to exclude such verbs is directly related to our selection of languages. In both classes of verbs we can expect massive convergence in Central Europe since all languages of the region have nominative-accusative alignment.[7] As has been said, however, this similarity is completely uninformative since this alignment type prevails not only in the region and not only in all of Europe. It is the most common type in the world (Nichols 1992 after Koptjevskaja-Tamm 2011, p. 579).

There is one more group of verbs that ought to be excluded from our choice of verbs. I will illustrate the problem with German examples.

(2) Ich   warte   auf meinen   Papa. (G, 5)
    1SG   wait.1SG   for my.M.ACC   dad.M.ACC
    'I'm waiting for my dad.'

---

7   Prevalence of the same alignment type among the languages of the region does not mean that the classes of plain intransitive and plain transitive verbs have exactly the same membership across these languages. They clearly do not, which leaves room for comparative studies. However, this topic is not pursued here.

(3) Ich   stelle   das   Glas   auf   den   Tisch. (G)
     1SG   put.1SG   the.N.ACC   glass   on   the.M.ACC   table
     'I'm putting the glass on the table (in a standing position).'

The last phrases in both (2) and (3) have the same marking, i.e., *auf* + ACC. However, the rules of the marking are different. The verb *warten* 'to wait', as in (2), consistently combines with *auf* + ACC for one of its arguments. By contrast, the verb in (3), *stellen* 'to put (in a standing position)', can combine with various spatial prepositions. The one used in a given sentence depends on the spatial configuration (produced by the action of putting) rather than on the verb.

Admittedly, languages do differ in how they encode spatial relations, which opens a huge and fascinating research field for comparative studies. Yet, this field falls outside of the scope of valency. While languages differ in the repertoire and exact meaning of markers of spatial relations, they are universally similar in that a locational argument of a verb allows a whole (sub)paradigm of markers available for marking spatial relations in the given language (Haspelmath/Hartmann 2015, p. 68).[8] Consequently, verbs that take typical locational arguments, like the one in (3), should be excluded from our data. It is not the variability of marking as such that is problematic (cf. the discussion of data from Tab. 2 below), but the fact that this variability follows from the general rules of a given language and not from the use of a particular verb as a valency carrier.

To sum up, there are at least three large groups of verbs which are of little value for our research agenda and should be excluded *a priori* from the verbal meanings selected for the data.[9] These are plain intransitives and plain transitives (i.e., verbs that only have arguments whose marking is predictable from the language's alignment type), as well as verbs with locational arguments (i.e., those which have arguments whose marking is variable in the same way as that of locational adjuncts). Note that exclusion of plain transitives does not mean exclusion of all nominative and accusative markers. These are found with numerous verbs that still have other arguments. The latter class of verbs is what this paper and my whole project focus on.

---

8   The said variability is found universally with all phrases with locational meaning, both arguments and adjuncts.

9   Note that these classes of verbs are not ideally matched among languages (cf. fn. 7). Thus, transforming a list of "uninformative" verbs into a list of verbal meanings to be excluded is a separate issue, which is discussed in Gaszewski (in preparation b).

Tab. 2: Realisations of sentence 2 from the input for data collection

sentence	language	marking for the			
		buyer	product	seller	price
Koupil jsem auto od souseda za milion.	Czech	(NOM)	ACC	od + GEN	za + ACC
Ich habe von meinem Nachbarn ein Auto für eine Million gekauft.	German	NOM	ACC	von + DAT	für + ACC
Vettem egy kocsit/autót a szomszédomtól egy millióért.	Hungarian	(NOM)	ACC	ABL	CAUSAL-FINAL
Kupiłem od sąsiada samochód za milion.	Polish	(NOM)	ACC	od + GEN	za + ACC
Am cumpărat o maşină de la vecinul meu pentru/cu un million.	Romanian	(NOM)	ACC	de la + ACC	pentru + ACC or cu + ACC

Note: NOM in brackets stands for the subject in the nominative not realised in the actual sentence.

2.3 The Essentials of Comparing Valency

If we were to compare valency in, say, Polish and English, the obvious counterpart for the Polish sentence in (1) would be the translation with *to buy* provided as a gloss in the example. The comparison must at least start with semantic equivalence. As we have said, semantics permeates valency frames and makes not only verbs, but also the valency positions associated with them directly comparable. Tab. 2 compares (1) and its equivalents in all the analysed languages.

Tab. 2 can be related to some of the issues raised above. For example, the pervasive nominative-accusative alignment is reflected in the first two columns. We can also clearly see the evident closeness of Czech and Polish—nearly all the words are recognizable as similar. This even applies to the adpositions marking valency, which are identical in form. Note also the availability of two markers in Romanian in the last column. The respective argument does not have a fixed marker in this language. This particular variation is not predictable from a cross-linguistic perspective, which is very different from locational arguments, cf. the discussion of (3). For this reason, marker variation like *pentru* + ACC/*cu* + ACC with the Romanian *a cumpăra* 'to buy' is indeed interesting from our perspective, unlike the marking of locational arguments.

Yet, the most important thing related to Tab. 2 is the mechanics of comparison. We put Czech *od* + GEN, German *von* + DAT, Hungarian ABL, etc. in one of the columns, because the respective phrases *od souseda*, *von meinem Nachbarn*, *a szomszédomtól*, etc. (all meaning 'from (my) neighbour') have identical roles in the valency frames constituted by the semantically equivalent verbs in the respective languages. In other words, these phrases have the same microrole associated with the verbs meaning 'to buy' and this is the exact foundation of the comparison.

Let us analyse the path of establishing equivalence of markers step by step. The sentences in Tab. 2 are semantically equivalent, the verbs are semantically equivalent, the respective sets of phrases have the same microroles and are semantically equivalent. Following upon this, we equate the cases and adpositions used as valency markers. An important caveat is that we only consider the markers in the relevant context. Whether the markers in question are equivalent elsewhere does not matter.

The same kind of comparison must be applied to other selected verbs, too. It is clear that the grammatical markers of valency in any language are much less numerous than valency carriers. Thus, as the comparison progresses, most of the markers are bound to resurface. What we are interested in is which microroles share marking. This is consistent with the approach advocated for comparative studies of valency in Hartmann/Haspelmath/Cysouw (2014) and used in Haspelmath/Hartmann (2015). Gaszewski (2012) somewhat intuitively employs a similar approach in the analysis.

Thus, the unit of comparison is a single equivalence of markers for a particular microrole. It is clear that some equivalences will be found in the data more frequently than others. The recurrent equivalences of valency markers reflect the patterns shared by the compared languages. The following are two sets of German–Czech equivalences.

(4) (German)   (Czech)
      warten + auf + ACC   ~   čekat + na + ACC
      'to wait for sb/sth'
      achten + auf + ACC   ~   dávat si pozor + na + ACC
      'to watch out for sb/sth'
      sich verlassen + auf + ACC ~   spoléhat se + na + ACC
      'to rely on sb/sth'
      schauen + auf + ACC   ~   koukat se + na + ACC
      'to look at sb/sth'

(5)   (German)   (Czech)
       betreffen + ACC   ~   týkat se + GEN
       'to concern sb/sth'
       entziehen + ACC   ~   zbavit + GEN
       'to deprive of sth'
       fragen + ACC   ~   zeptat se + GEN
       'to ask sb' (pose a question)

Let us stress again that the markers in (4–5) are equivalent in the given contexts, which is distinct from general semantic equivalence. The two kinds of equivalence may well coincide as they do in the case of (4), since German *auf* + ACC regularly corresponds to Czech *na* + ACC. In (5), by contrast, the two kinds of equivalence do not coincide.

On the most abstract level of comparison, the selected verbs and the microroles associated with them create an abstract semantic space in which the cases and adpositions are distributed. The patterns of their distribution are the immediate object of the present analysis. To reformulate the main objective (cf. section 1), we want to know how parallel the patterns of distribution of valency markers are among the analysed languages of Central Europe.

It is of crucial importance that the abstract semantic space in question is populated by valency markers of each language separately. Following Haspelmath (2010), I treat every grammatical marker of any language as essentially separate from anything in another language. Each marker is a *descriptive category* within its particular language system. This does not mean that languages cannot be compared or that there are no similarities in the internal structuring of the systems. Rather, nothing is really *the same* between grammars of different languages. In the context of the present study it means that the markers are all different across languages, however, identical equivalences recur, as shown by (4–5). In practice, the full description of a marker must therefore name the language. For example, the markers of the arguments in (4) are: G_auf + ACC and Cz_na + ACC.

To illustrate this approach, let us consider the cases labelled accusative in the analysed languages (i.e., Cz_ACC, G_ACC etc.). These markers appear as equivalents in Tab. 2 and such equivalences are very common in our data (cf. section 3.3). However, this observation is an empirical outcome of the comparison. Even though it is fully predictable (cf. section 2.2), at no point do we need to assume that these cases are instantiations of some universal accusative category. Haspelmath's (2010, p. 666) point here is that even when markers like these coincide across languages, they are not identical. This is reflected in our data— no single language pair has an ideal match of accusatives.

3 Methodology

3.1 Data Collection

The point of departure for data collection was the list of selected verbal meanings. Because of the complexity of the issue (cf. section 1), the principles of the selection are described in a separate paper (Gaszewski in preparation b). Here, I elaborate on procedures that build upon a ready list. The data collection itself consisted of three steps. First, example sentences were created for the selected verbal meanings. Each was meant to instantiate the valency frame associated with a given verbal meaning. The very sentences were in English, which is the best practical choice as the modern-day lingua franca, but has clear drawbacks too.

In the second step, the sentences were used as input to collect the data. The informants were given the sentences and had to translate them into their native languages. They were asked to focus on the naturalness of their output and had the opportunity to add comments with qualifications and clarifications as they saw fit. Tab. 3 shows the format of presentation of the sentences with the data filled in by the Romanian informant.

The informants not only had to provide the full-sentence translation, but also match parts of the original and translation. For each part of the sentence, both the exact form used in the sentence and the *basic form* were to be provided. The instruction given to the informants was that the basic form should include forms that one can find as dictionary entries.[10] The purpose of having the two variants of the same part of the sentence was to facilitate the identification of the relevant grammatical markers in the data.

The last step was the annotation of the data, done by the author. The grammatical forms used in the sentences (cases and adpositions) were noted down separately for each valency position of each verb in the collected data. Tab. 4 presents the effects of annotation for some of the structures, mostly included in the examples above. A single line in Tab. 4 constitutes one datapoint.

We must still note three ways in which data collection allowed for variation in the data. First of all, it may be the case that a language allows several grammatical forms to mark a single microrole. This has already been mentioned with

---

10 The label *dictionary form* as such was shunned because very often the given part of the sentence consisted of several words and the ideal basic form was a corresponding phrase in the nominative which would not be directly found in a dictionary (cf. last line of data in Tab. 3).

Tab. 3: Input for data collection with Romanian data filled in (table for sentence 14)

SENTENCE NUMBER	14
ORIGINAL SENTENCE	**This belongs to my mother.**
TRANSLATED SENTENCE	*Asta aparține mamei mele.*
ELEMENT CORRESPONDING TO 'belongs'	*aparține*
BASIC FORM OF THIS ELEMENT	*aparține*
ELEMENT CORRESPONDING TO 'to my mother'	*mamei mele*
BASIC FORM OF THIS ELEMENT	*mama mea*
NOTES	

Note: The input is shown in roman font, data provided by the informant is in italics.

Tab. 4: Example lines of annotated data

	language	verbal meaning	microrole	valency carrier	grammatical marker
1	Polish	BUY	BUY-1	kupować	Pl_NOM
2	Polish	BUY	BUY-2	kupować	Pl_ACC
3	Polish	BUY	BUY-3	kupować	Pl_od + GEN
4	Polish	BUY	BUY-4	kupować	Pl_za + ACC
5	Romanian	BUY	BUY-4	a cumpăra	R_pentru + ACC
6	Romanian	BUY	BUY-4	a cumpăra	R_cu + ACC
7	Romanian	BELONG	BELONG-2	a aparține	R_DAT
8	German	WAIT	WAIT-2	warten	G_auf + ACC
9	Czech	WAIT	WAIT-2	čekat	Cz_na + ACC
10	German	WATCH-OUT	WATCH-OUT-2	achten	G_auf + ACC
11	Czech	WATCH-OUT	WATCH-OUT-2	dávat si pozor	Cz_na + ACC
12	Hungarian	DEPEND	DEPEND-2	függ	H_ABL
13	Hungarian	DEPEND	DEPEND-2	múlik	H_SUPERESS

reference to Romanian data in Tab. 2. The variant markers are shown in Tab. 4 in lines 5 and 6.

Secondly, note the Czech phrase *dávat si pozor* 'to watch out, to be careful' in (4) and included also in Tab. 4 (line 11). The phrase is not strictly a verb but a light verb construction, but such variation of the formal category of the predicate was allowed, as long as the predicate represented the verbal meaning in question.

Thirdly, lines 12 and 13 in Tab. 4 show two Hungarian verbs that were provided for a single sentence of the input. Synonymy of lexical items is a phenomenon

present in languages and should not be ignored. Hungarian *függ* and *múlik* were accepted as separate realisations of the same verbal meaning. Consequently, their arguments bear the same microroles. Note well that synonymous verbs may not require the same morphological marking, which is the case with the two Hungarian verbs.

3.2 Data Analysis

The collected data comprise several hundred datapoints, which makes some form of descriptive statistical modelling necessary for adequate interpretation of the data. The method described here focuses on establishing distances between the languages.

The idea of treating the markers of each language as separate (cf. the final part of section 2.3) restricts the possible format in which the data can be analysed. The simplest tabulation possible, in which microroles are set against languages (e.g., in Tab. 2) proves sufficient for illustration, but unproductive for statistics. To be used for calculations, the table requires single values in all slots. Thus, it cannot properly account for observed marker variation shown in Tab. 2 and 4. Moreover, the values in such tables remain nominal, the lowest measurement level from the point of view of statistics. The alternative is to tabulate languages pairwise against each other. The new format provides numerical data. Tab. 5 below is a realisation of that model for Czech and Hungarian data.

The slots in the table represent all logical possible equivalences of Czech and Hungarian markers in the data.[11] The number in each slot says how many times the equivalence of the Czech marker (for the given row) and the Hungarian marker (for the given column) occurs in the dataset. Most of the slots have zeros—the corresponding equivalences are not attested. The rows and columns are both sorted according to the marginal sums so that the bottom left corner is populated by the highest values. Tab. 5 is not easily legible, but it is a set of numbers for calculations. Fig. 1 provides a more reader-friendly illustration of the data sorted the same way. The sizes of the dots correspond to the numbers in the slots of Tab. 4. The zeros, shown as lack of a dot, are more conspicuous.

The numbers from a table of this kind can be used to determine how similarly or differently the two languages distribute their valency markers. Languages

---

11 Grammatical markers totally absent from the data (e.g., the Hungarian adessive case) are not included in the table. The practical result is that every row and column has a marginal sum larger than 0.

Tab. 5: Contingency table for equivalences of Czech and Hungarian markers

	Hungarian							
Czech	H_accusative	H_ dative	H_ sublative	H_ ablative	H_causal-final	H_ delative	H_ instrumental	H_nominative
Cz_z + genitive	0	1	0	0	0	0	0	0
Cz_v + locative	0	0	0	0	0	0	0	0
Cz_v + accusative	0	0	0	0	0	0	0	0
Cz_proti + dative	0	0	0	0	0	0	0	0
Cz_pro + accusative	0	0	1	0	0	0	0	0
Cz_před + instrumental	0	0	0	1	0	0	0	0
Cz_jako + nominative	0	1	0	0	0	0	0	0
Cz_o + accusative	1	0	0	0	1	0	0	1
Cz_nominative	1	0	0	0	0	0	0	2
Cz_na + locative	0	0	0	1	0	0	0	1
Cz_k + dative	1	0	1	0	0	0	0	0
Cz_instrumental	0	1	0	0	0	0	1	0
Cz_do + genitive	0	0	2	0	0	0	0	0
Cz_od + genitive	0	0	0	4	0	0	0	0
Cz_genitive	2	1	0	2	0	0	0	0
Cz_s + instrumental	1	0	1	0	0	0	4	0
Cz_o + locative	3	0	0	0	0	5	0	0
Cz_za + accusative	1	1	1	0	7	0	0	0
Cz_na + accusative	4	0	4	0	0	2	0	0
Cz_dative	3	16	0	0	0	0	0	0
Cz_accusative	33	0	5	0	0	0	1	1
TOTAL	**50**	**21**	**15**	**8**	**8**	**7**	**6**	**5**

H_ inessive	H_ allative	H_ nominative + ellen	H_ translative	H_ nominative + elől	H_ illative	H_ nominative + mellett	H_mint + nominative	H_ superessive	TOTAL
0	0	0	0	0	0	0	0	0	**1**
2	0	0	0	0	0	0	0	0	**2**
1	0	0	1	0	0	0	0	0	**2**
0	0	2	0	0	0	0	0	0	**2**
0	0	0	0	0	0	1	0	0	**2**
0	0	0	0	1	0	0	0	0	**2**
0	0	0	0	0	0	0	1	0	**2**
0	0	0	0	0	0	0	0	0	**3**
0	0	0	0	0	0	0	0	0	**3**
0	0	0	0	0	0	0	0	1	**3**
0	1	0	0	0	0	0	0	0	**3**
0	0	0	1	0	0	0	0	0	**3**
0	0	0	0	0	1	0	0	0	**3**
0	0	0	0	0	0	0	0	0	**4**
0	0	0	0	0	0	0	0	0	**5**
1	0	0	0	0	0	0	0	0	**7**
0	0	0	0	0	0	0	0	0	**8**
0	0	0	0	0	0	0	0	0	**10**
0	0	0	0	0	0	0	0	0	**10**
0	1	0	0	0	0	0	0	0	**20**
0	0	0	0	0	0	0	0	0	**40**
4	**2**	**2**	**2**	**1**	**1**	**1**	**1**	**1**	**135**
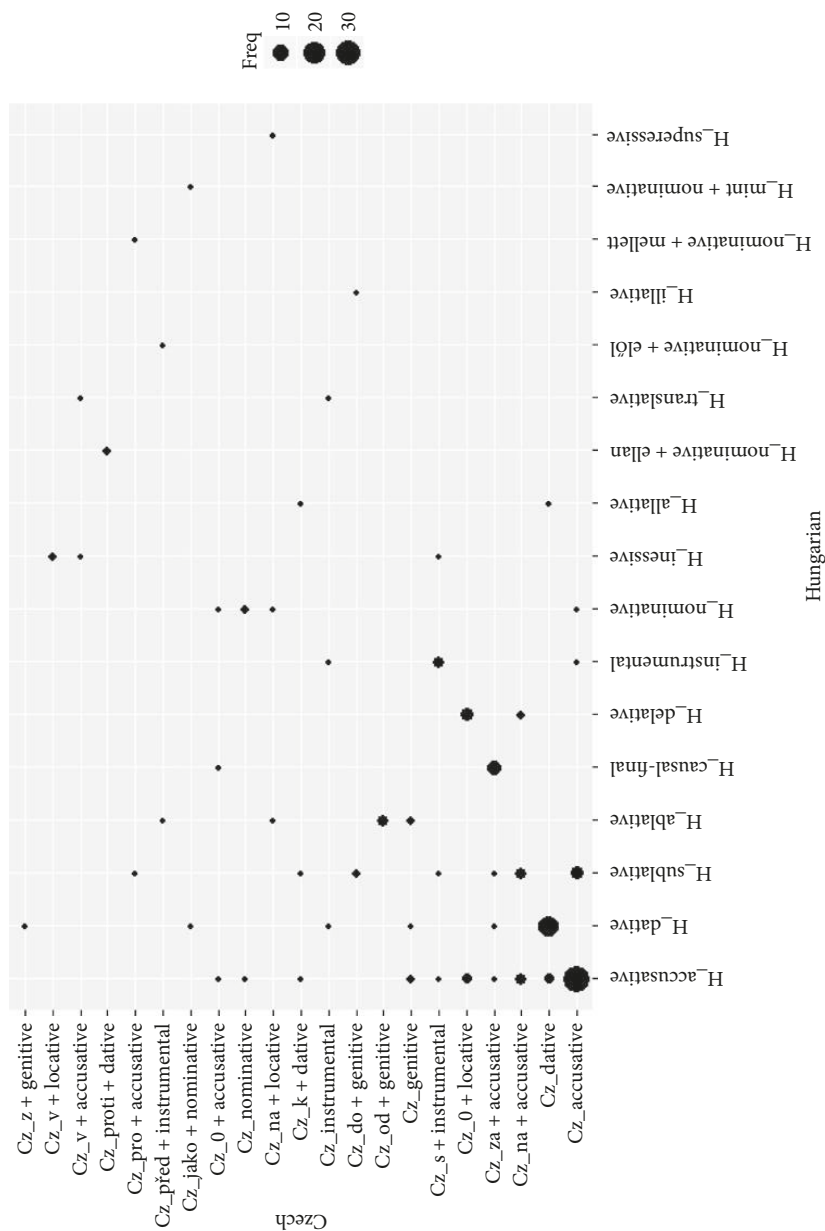
Fig. 1: Illustration of the distribution of equivalences of Czech and Hungarian markers (distance = 0.510)
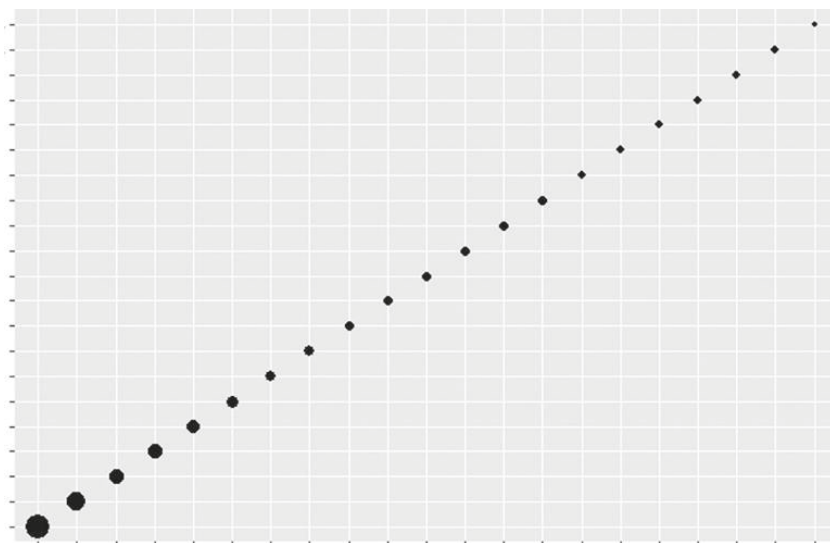
Fig. 2: Illustration of the ideally matched distribution of markers (distance = 0). The figure was produced by tabulating data from one language against itself.

with ideally matching markers would produce a table with even fewer non-zero slots—exactly one per column and one per row (the corresponding graph is shown in Fig. 2).

Evidently, no language pair in our data gives such a picture. However, we can use a distance index, a simple statistical measure, to establish how much each pair diverges from the ideal match. In other words, we calculate how distant the two languages are from each other regarding the distribution of grammatical markers.

For this purpose, in each column a single highest (modal) numeric value is found. Modal values from all columns are summed, the remaining (non-modal) values from all columns are also summed. The resultant non-modal sum is divided by the corresponding modal sum. The same procedure is done row by row. The distance index is the average of the quotients for columns and rows.

The slots with modal values (i.e., those with the greatest numbers of recurring equivalences of markers) lower the value of the distance index and stand for the similarity in the distribution of valency markers. By contrast, the slots with non-modal values raise the value of the distance index and represent the scattering

Jerzy Gaszewski
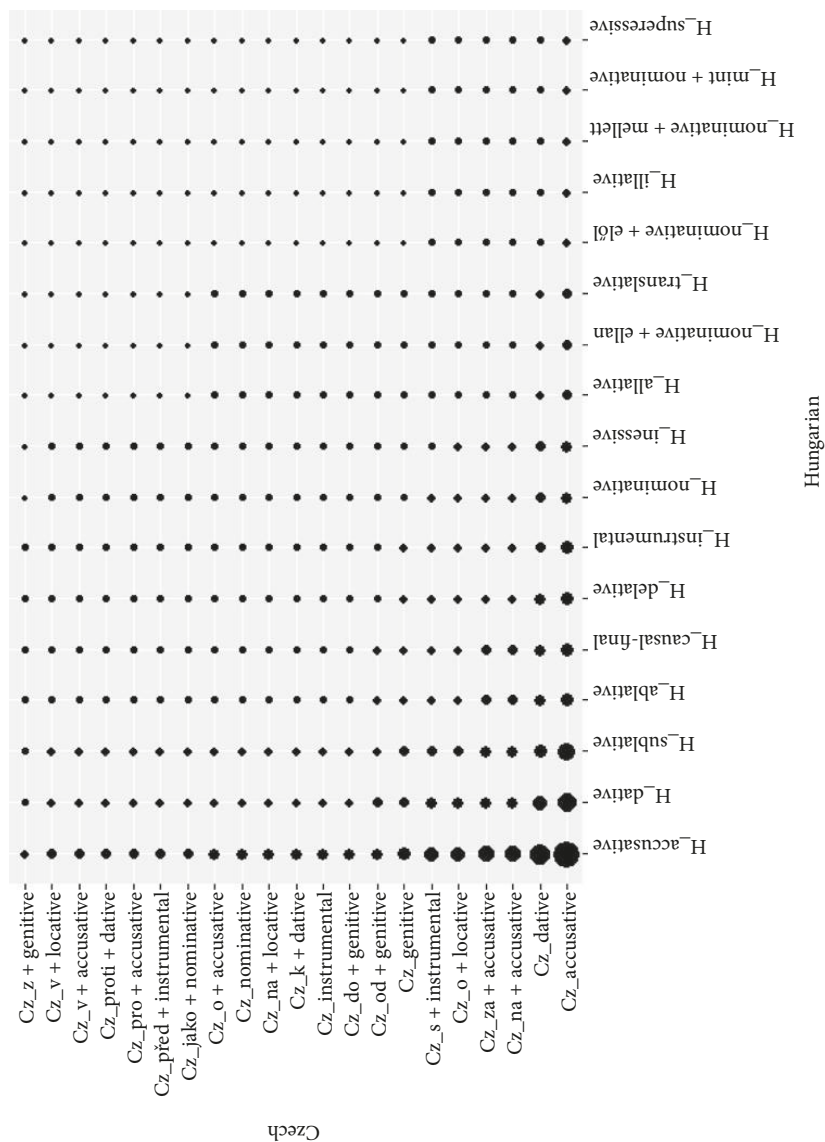
Fig. 3: Even spread of data based on marginal sums of Czech and Hungarian data (distance = 2.038)

of the data and the dissimilarity of the distribution. A distance index value of 1 corresponds to equal sums of non-modal and modal slots.

There is no upper limit for the distance index as such. However, we can calculate the maximal possible distance for each set of data. The maximal distance corresponds to the even numerical spread within the table and can be derived from the marginal sums of a given actual dataset.[12] The even spread (illustrated in Fig. 3) is the opposite of the ideal match.

As can be predicted, actual data give results that fall somewhere between the two extremes. The analysis of the data (cf. section 4.1) consists in calculating and comparing the distances for all language pairs (five languages yield ten pairs).

3.3 The Proper Dataset

Having decided on a method of analysis, we must still make an equally important decision about the most appropriate subset of the data to which to apply the method. As section 2.2 showed, the markers associated with regular transitive clauses are problematic for our research agenda. This is reflected in the collected data. Cases called nominative[13] in the analysed languages massively correspond to one another across the data. The same effect can be observed for the cases called accusative, albeit to a lesser degree.

The nominative and accusative markers cover a substantial portion of the microroles, which means that the similarity produced by equivalences of these markers will have an impact on the distance indices. One problem is that the similarity is completely predictable from the alignment type. Another problem is that it is shared by all the compared language pairs. Thus, as long as all the analysed languages pattern the same way, the nominative and accusative markers found in the data are uninformative and should not be taken into account in the calculations.[14]

---

12   The values corresponding to the even spread of data, called *expected* values, are calculated in the following way: for each slot the sum of the row is multiplied by the sum of the column and then divided by the grand total.

13   It is not the repetition of the grammatical labels itself that is problematic. Rather, the matching terminology reflects the factual general match of nominative and accusative cases across languages.

14   This is a logical continuation of the removal of plain intransitive and transitive verbs from the analysis (cf. section 2.2). These are all verbs whose arguments have predictable marking. Now we decide to remove the arguments with predictable marking of verbs that also have arguments with other markers.

Note the requirement that the transitive marking must be shared by all the analysed languages to be excluded from further analysis.[15] If this is the case, the observed similarity can surely be treated as background that brings nothing into the comparison and the corresponding microroles should be removed from the counts. However, if the marking for a given microrole diverges from the shared pattern in a single language, that microrole remains within the analysed dataset because it differentiates this language from the rest.

The following figures show how the exclusion of the all-matching nominative and accusative markers relates to the distribution of valency markers and how it affects the value of the distance index. The figures refer to the German-Polish pair. Fig. 4 takes into account all the microroles in the collected data for the two languages. Fig. 5 does not include all-matching nominatives, but has all-matching accusatives.[16] There is a notable rise in the distance index. However, when we compare the distribution of dots in the two figures, very little has changed. Indeed, Fig. 5 is basically Fig. 4 without the one largest dot for the nominative-nominative equivalence.

Fig. 6 renders the distribution of markers with both all-matching nominatives and all-matching accusatives removed from the count. Again, there is a tangible rise in the distance index, but the change in the distribution of all the markers is slight. Unlike in the case of nominatives, which virtually disappear between Fig. 4 and Fig. 5, a considerable portion of the accusative markers remains in Fig. 6. There is still a dot for the accusative-accusative equivalence shared by German and Polish, but not all the five languages.

The figures prove that the all-matching nominative and accusative markers do affect the value of the distance index (more than twice as high for Fig. 6 than for Fig. 4). When they are removed, the value of the distance index is based on the other markers, whose distribution is not predictable from the alignment type of the languages. In this way, thanks to the removal of part of the data, the analysis focuses on the markers that may realistically differ.

All the data collected in the course of the pilot study includes 192 microroles and 1016 datapoints. The dataset used for calculations comprises 96 microroles and 521 datapoints in total. The excluded microroles are marked in Appendix 1.

---

15 The requirement relates to a particular set of data and languages. In general, we can expect that addition of more languages to the comparison might reduce the group of microroles removed from analysis.

16 This level of restriction on the data also applies to the figures in section 3.2.
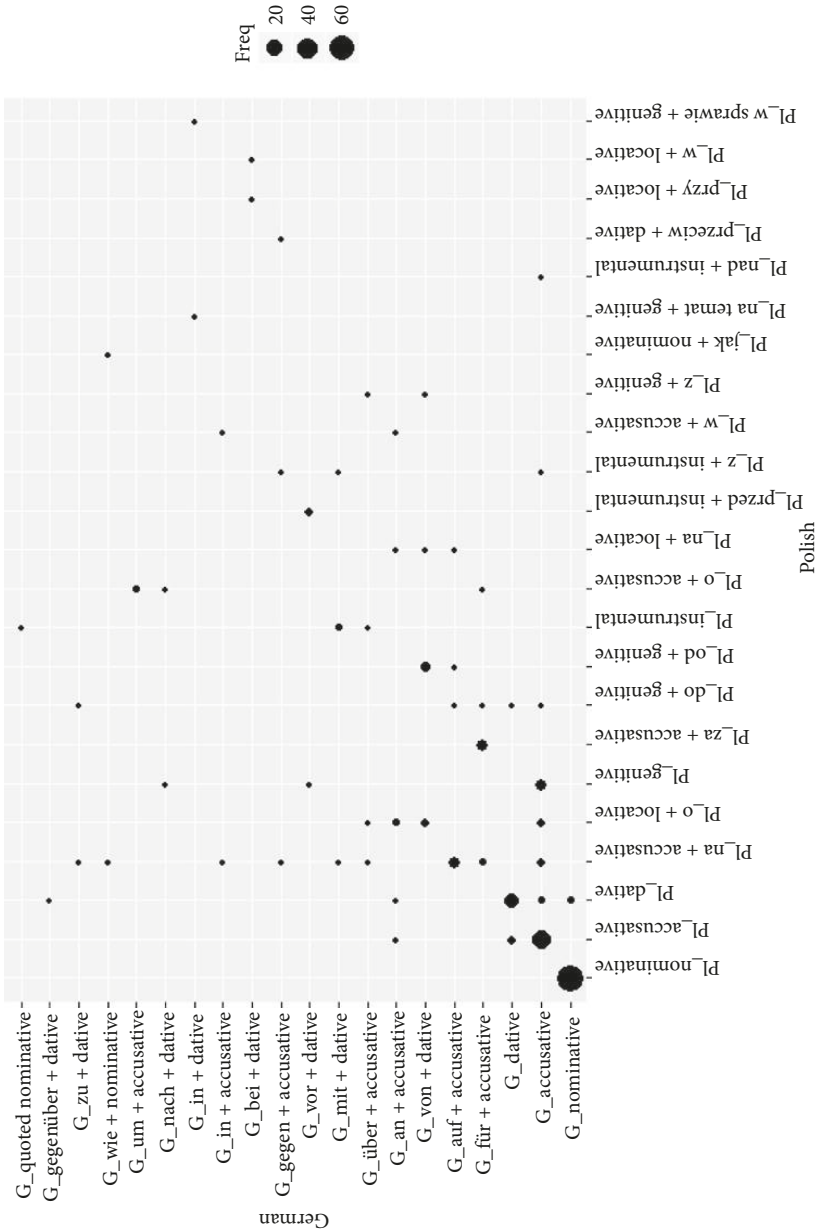
Fig. 4: Distribution of German and Polish markers for all microroles in the collected data (distance = 0.309)
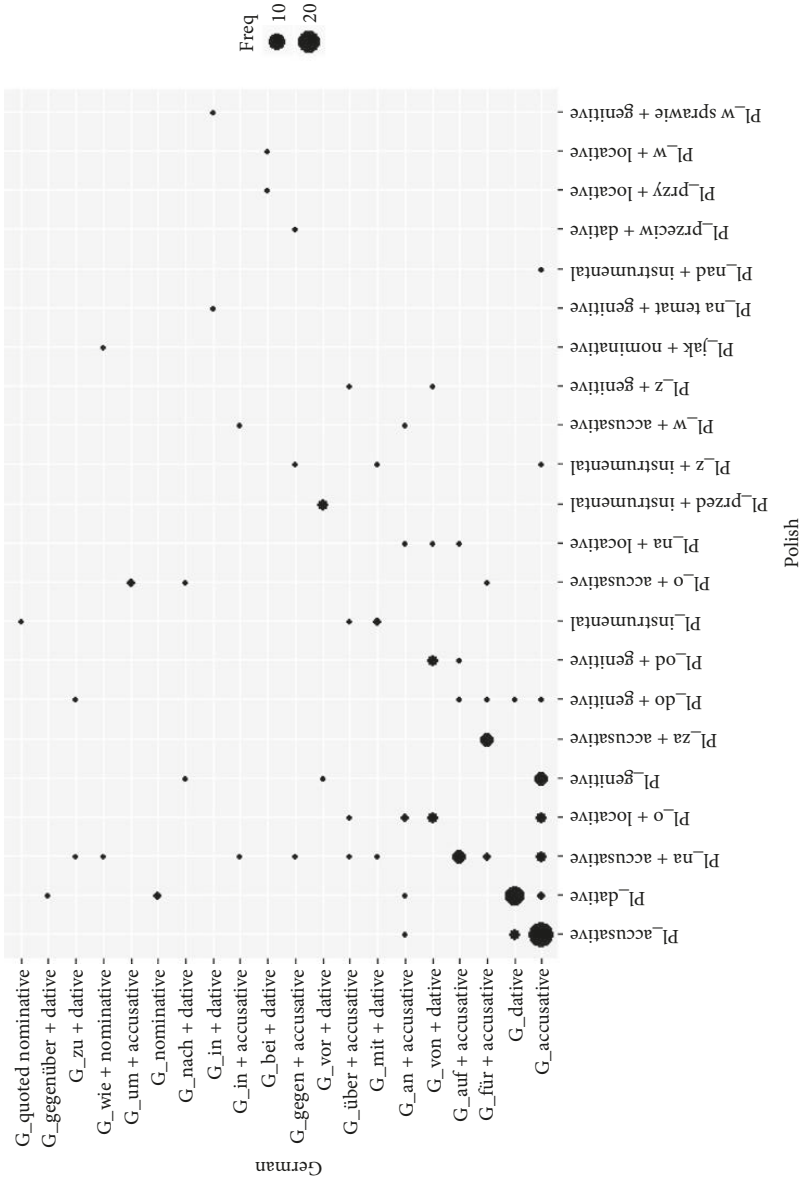
Fig. 5: Distribution of Polish and German markers without microroles with all-matching nominatives (distance = 0.565)
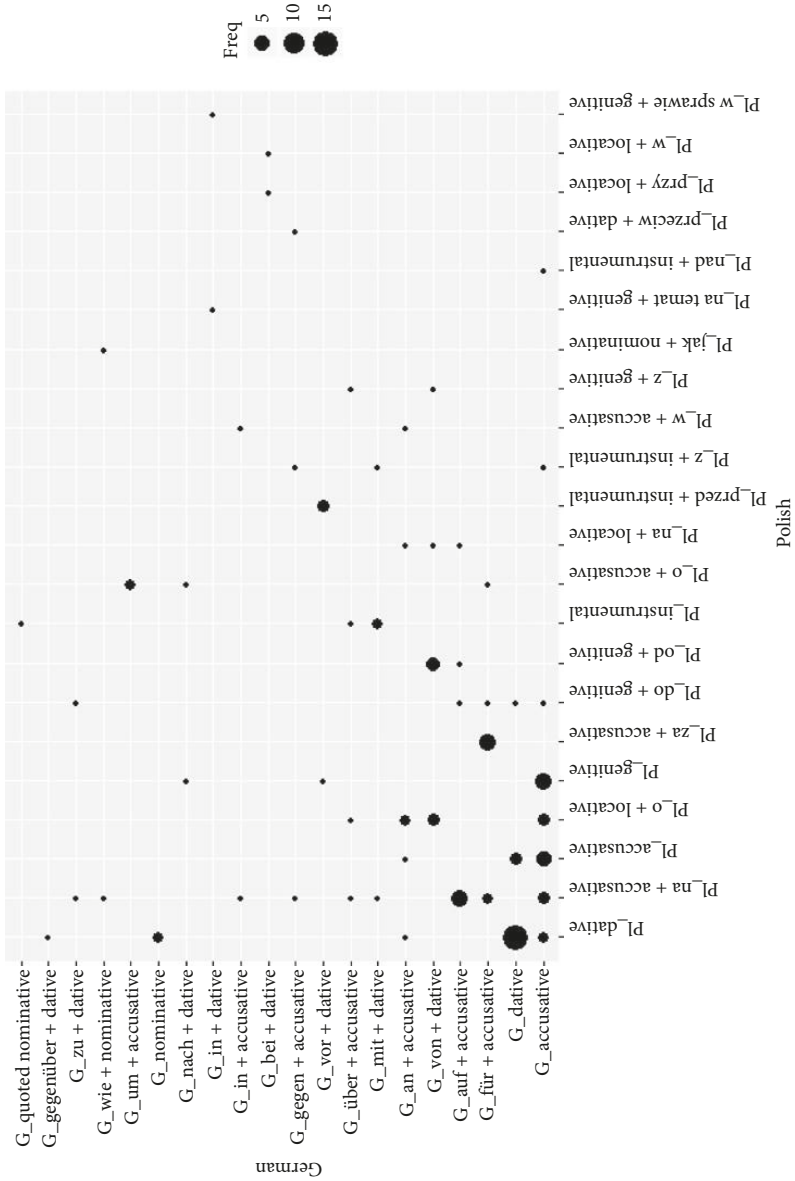
Fig. 6: Distribution of Polish and German markers without all-matching nominatives and accusatives (distance = 0.775)

3.4 The Correction for Cognates

As we have noted in the beginning, there are two closely related languages in our group, Czech and Polish. The results (cf. Tab. 6 below) confirm the similarity of the two languages. However, genetic affinity is a confounding factor for an areal study. We cannot tell *a priori* how much of the similarity is due to common origin. In order to solve the problem, I propose a correction for cognates in this language pair.

Let us now examine what the said genetic similarity means in practice.

(6)   (Czech)   (Polish)
       čekat + na + ACC   ~   czekać + na + ACC
       'to wait for sb/sth'

The verbs in (6) are clearly reflexes of a single Common Slavic form **čekati*. The prepositions are also cognate, they even have exactly the same form. They both combine with accusative case forms. The Czech and Polish accusatives may have different forms of the endings for some words, but they certainly are continuations of the same Common Slavic category. In fact, the same can be said about the whole case systems of the two languages (cf. Comrie 2009, p. 273).

In light of the above, an equivalence like (6) appears very suspect from the areal perspective. It is very likely that it represents the genetic similarity of the two languages. We are speaking of likelihood here because the assessment of (6) is based on the cognate status of parts of the construction, which does not entail that the whole structure is indeed inherited from a common earlier stage of development by both languages. One can attempt to establish whether this is the case by analysing extant texts, corpora, etc. Yet, it is beyond our means to do this for all potential cognate constructions. Instead, we apply the correction for cognates, which is a heuristic measure.

Further examples illustrate other situations encountered while assessing potential cognate structures.

(7)   (Czech)   (Polish)
       vyměnit + za + ACC   ~   wymienić + na + ACC
       'to exchange for sth'

While the verbs in (7) are again cognates, we have etymologically distinct prepositions. This is a crucial difference in comparison to (6) since our focus is exactly on the distribution of the grammatical markers. We cannot treat the two equivalences in the same way. While (6) is removed from the data by the correction for cognates, (7) must necessarily remain.

The correction relies on the cognate status of the elements of the constructions in question. In practice, it is not always clear what exactly qualifies as cognate.

(8) (Czech) (Polish)
ukázat + DAT ~ pokazać + DAT
'to show to sb'

In (8) we have a pair of verbs which obviously share the same root (Proto-Slavic *_kazati_), but have different prefixes, i.e., they are not fully parallel etymologically. One could argue both for including and excluding such datapoints. The correction for cognates involves reduction of the data, which in itself is not very desirable. Thus, I decided to be rigid in the understanding of what defines a cognate and to keep partial cognate equivalences like (8) in the data.

To sum up, the correction for cognates means removing from the calculation all equivalences of valency structures where _both_ the verb and the respective marker are cognate in Czech and Polish. The recalculated distance is of course higher than the one based on Czech and Polish data without the correction. It is also important that the Czech–Polish dataset after the correction is reduced in comparison to other language pairs. Thus, the correction is an accessory measure and the results it produces should be presented together with the non-corrected ones.

The same procedure could in principle be applied to the other language pairs as well. Although Czech and Polish are the closest pair by far, both being West Slavic languages, Romanian and German are also distantly related to them, and to each other. However, the number of cognate structures that would be found is negligible and thus only the Czech–Polish pair underwent the correction for cognates. Yet, when we proceed to enlarge the group of the analysed languages, there will be more pairs of closely related languages and the correction will have to be applied to them too.

4 Results and Conclusions

4.1 Distances between the Languages

Distance indices for each language pair were calculated following the procedure described in section 3.2, with the restrictions on the dataset described in section 3.3. The values for all pairs are shown in Tab. 6, ranked from the lowest distance index (most similar distribution of valency markers) to the highest (most divergent distribution). For the Czech–Polish pair there are two results: without and with the correction for cognates (cf. section 3.4).

Tab. 6: Ranking of language pairs according to distance index

language pairs	distance index	language pairs with the correction for cognates
Czech vs. Polish	0.490	
Czech vs. German	0.562	Czech vs. German
German vs. Hungarian	0.666	German vs. Hungarian
Czech vs. Hungarian	0.687	Czech vs. Hungarian
	0.697	*Czech vs. Polish*
German vs. Romanian	0.715	German vs. Romanian
Czech vs. Romanian	0.717	Czech vs. Romanian
Hungarian vs. Romanian	0.769	Hungarian vs. Romanian
German vs. Polish	0.775	German vs. Polish
Hungarian vs. Polish	0.805	Hungarian vs. Polish
Polish vs. Romanian	0.816	Polish vs. Romanian

Tab. 7: Ranking of languages according to mean distances

languages	mean of distance indices	languages with the correction for cognates
Czech	0.614	
	0.666	*Czech*
German	0.680	German
Polish	0.721	
Hungarian	0.732	Hungarian
Romanian	0.755	Romanian
	0.773	*Polish*

Tab. 7 displays a ranking of languages on the basis of the mean of distance indices for the pairs in which a given language was involved. The lower the mean, the closer a language is to all the others (and the higher it is ranked in the table).

Occupying the close top positions in Tab. 7, Czech and German prove to be the most central languages in the group. This corresponds well with the fact that this pair ranks first in Tab. 6 (with the correction for cognates). Undoubtedly, areal convergence must be at work here, given the geographical closeness and long history of mutual contact, without close genetic affinity.

Hungarian ranks third in Tab. 7 (with the correction). This is an interestingly good position for the only language that is not related to the others. We can rather firmly say that at least in terms of distribution of verb valency markers,

Hungarian does not stand out in the region, which again suggests the importance of areal patterning. Note also that the second and third closest pairs in Tab. 6 (with the correction) involve Hungarian and the central languages Czech and German.

Romanian is fourth in Tab. 7 (with the correction). This language is generally considered to belong to the Balkan sprachbund. Yet, Romanian fits our group well enough, and its expected status of a "natural outlier" (cf. section 1) is not confirmed by the data.

Polish was pushed to the last position in Tab. 7 by the correction for cognates (from the middle of the ranking). Most surprisingly the language is finally positioned below Romanian. A close inspection of Tab. 6 shows that this result does match other numbers well. The Czech–Polish pair ranks high (top without the correction for cognates, and middle with the correction), but the three other pairs involving Polish are blocked at the bottom of the ranking. The pairs involving Romanian (apart from the Polish–Romanian one) rank higher. In sum, the data do show Polish to be the most dissimilar language in the group.

The correction for cognates affects the values of Czech and Polish. However, only for Polish does this lead to a drop in the ranking in Tab. 7, and a notable one at that. Evidently, genetic similarity did push this language upwards in the ranking without the correction. In total, about 26 % of the Czech–Polish data was removed in the process. This is certainly a substantial part, but the reduction is not overwhelming. The genetic closeness of the two languages does not overshadow their distinct identity, at least as far as valency marking is concerned.

Thus far, we have focused on the rankings, i.e., how the languages are positioned relative to one another. We shall now consider the actual values of the distances. Note that the distance index has been introduced as a new measure here and the interpretation of the values is not obvious. However, we can propose some points of reference that can help in the following discussion.

In section 3.3 it was decided that much of the predictably similar material must be excluded from the calculations. Yet, the values of distance indices in all language pairs still seem rather low as well as close to each other. One point of reference against which we can check these impressions are the maximal possible values of the distance index that correspond to the even spread of equivalences (cf. Fig. 3). The maximal possible distance indices for our data range from 3.077 (German vs. Hungarian) to 4.856 (Czech vs. Polish). These values are much higher than the actual distance indices in Tab. 6, which are only about 20 % of the respective maximal possible distances. Thus, we can confirm that the actual distances are low and much closer to the pole of ideal similarity than to the even spread of data.

A more rigid point of reference is the distance index value of 1, which corresponds to exactly equal numbers of modal and non-modal equivalences in a language pair. None of the language pairs has so high a value of the distance index. In other words, in each language pair the equivalences of markers supporting the similarity outnumber the ones supporting the dissimilarity. This shows the general similarity of the five languages,[17] which may be related to areal convergence. On the other hand, the above considerations about how the languages rank are put into perspective. The general similarity of all the analysed languages also raises the suspicion that it might be shared by more languages than just these five.

4.2 Conclusions

The results of the analysis presented in the previous section fit, to a considerable extent, the description of the selected languages in section 1. Czech and German are the most central languages in the group. Hungarian ranks third. The "core" of the group formed by these three languages corresponds well with the established linguistic landscape of the region, especially the *Donaubund* group. However, several elements spoil the picture.

The most important problem is that the distance indices in all pairs are rather low and have quite similar values. All the analysed languages are fairly similar to one another in the distribution of valency markers. In other words, we do not have very strong foundations for the firmness of the ranking of languages. The picture we have may of course be strengthened by further analysis with more data and more languages. Yet, it is very likely that an extension of the dataset will alter the picture, it remains to be seen how significantly.

Moreover, since all the languages seem to be quite similar overall, any talk of the "core" of the group, let alone "outliers" is rhetorical rather than factual, at least at the current stage. Still, let us examine how the values of the means in Tab. 7 progress. The only possible cut-off point is between German and Hungarian, where the difference between the means is by far the largest. If we can attach any meaning to this, it proves the data allows us to separate only the Czech–German pair as the putative "core". This does not support the *Donaubund* interpretation so much, but the unique closeness of these two languages only (in congruence with the ideas of Skála 1992 after Newerkla 2000, p. 2).

---

17 This should be taken with caution. We still have too little knowledge of how the distance index functions to assess the strength of the general similarity with certainty.

As has been said, the relative positions of Polish and Romanian are unexpected. The current dataset does not reveal any clear division between the Balkan languages (represented by Romanian) and Central Europe. A much less well acknowledged division along the northern Carpathian Mountains (separating Polish from the *Donaubund* proper) is at least as strong in the analysed dataset.

Despite all the reservations, we can say that our dataset gives some support for the areal grouping of *Donaubund*. A group of five elements can produce numerous permutations and still the order of the ranking in Tab. 7 is generally in line with the expectations. On the other hand, our small selection of five languages does not feature any other well-established group that could be expected to be a likely alternative. This calls very strongly for the investigation of a wider context, i.e., more languages. At this point we can only speculate how well the concept of *Donaubund* would fare with a larger set of languages. Note that the results we have may be supportive of this areal grouping, but they are consistent with larger areal groupings as well, e.g., the European linguistic area.

The expansion of the analysed group of languages has been mentioned since the very beginning as a necessary improvement in further research. Section 4.3 discusses all the extensions of the research agenda which follow from the shortcomings inevitable in a pilot study.

4.3 Outlook

Further work on verb valency in Central Europe must encompass a number of improvements. First, more languages must be included. It is worth considering *what* languages are to be added. The first idea might be to include more languages of the more broadly understood Central Europe (e.g., Croatian, Slovak, Ukrainian or Lithuanian). However, what we lack most are languages clearly outside of the area. These are very informative because they give a perspective against which the results referring to the languages within the region can be properly interpreted. Also, without such a broader areal background we cannot properly distinguish features characteristic of the region from features shared with other regions (cf. Haspelmath 2001). Thus, the ultimate group of languages must include a subset of what can be called *control languages*. The eventual choice is to use languages from geographical peripheries of Europe for this function, e.g., Spanish, Greek, Russian or Swedish. The data will be easy enough to collect, while the distances can be expected to be higher when these languages are paired with the Central European ones or with one another. These language pairs will provide a good data-driven point of reference against which the distance index values can be compared.

The input used for collecting the data also requires improvements. So far it has been the rule that one verbal meaning is represented by one example sentence given to one native speaker of each language. The number of verbal meanings, the number of example sentences as well as the number of informants should all be increased. These are the easy changes, at least conceptually. A more fundamental question is whether translation of sentences is sufficient as a method of data collection. The problems include potential transfer from English (although the informants are given clear instruction to aim for naturalness in their language), but also difficulty of expressing some structures in the input. It seems we need to combine sentence translation with another mode of data collection that would give more freedom to the informants. The idea of how to collect the data is described in Gaszewski (in preparation a).

Finally, the results presented here do not tell us everything we might want to know even about the relatively small dataset of the pilot study. We have the general picture of how the selected languages relate to one another as whole blocks of data. However, nothing has been said about the particular structures that account for the observed similarities and dissimilarities. It is worth examining which markers are involved in the frequently recurring equivalences and to compare language pairs in this respect (cf. Gaszewski in preparation c).

Appendix 1

All sentences used as input for collecting data in the pilot study are provided here. Microroles present in each sentence are listed beneath it. Whenever the marking of a microrole involves all-matching nominative or accusative cases in the analysed languages, a note is made.

(1) I'm afraid of ghosts.

| microroles: | FEAR-1 | 'person feeling fear' (all-matching nominatives) |
| | FEAR-2 | 'danger feared' |

(2) I bought a car from my neighbour for a million.

microroles:	BUY-1	'buyer' (all-matching nominatives)
	BUY-2	'product bought' (all-matching accusatives)
	BUY-3	'seller'
	BUY-4	'price'

(3) I took a book from Henry.

microroles:	TAKE-1	'person taking' (all-matching nominatives)
	TAKE-2	'thing taken' (all-matching accusatives)
	TAKE-3	'person from whom something is taken'

(4) I'm happy about the prize.

| microroles: | HAPPY-1 | 'person feeling happy' (all-matching nominatives) |
| | HAPPY-2 | 'thing rejoiced over' |

(5) I'm waiting for my dad.

| microroles: | WAIT-1 | 'person waiting' (all-matching nominatives) |
| | WAIT-2 | 'thing awaited' |

(6) I gave my mum a present.

microroles:	GIVE-1	'giver' (all-matching nominatives)
	GIVE-2	'recipient'
	GIVE-3	'thing given' (all-matching accusatives)

(7) I got a present from my friend.

microroles:	RECEIVE-1	'recipient' (all-matching nominatives)
	RECEIVE-2	'thing received' (all-matching accusatives)
	RECEIVE-3	'giver'

(8) It concerns my parents directly.

| microroles: | CONCERN-1 | 'issue concerning sb' (all-matching nominatives) |
| | CONCERN-2 | 'person concerned' |

(9) I thanked those people for their help.

microroles:	THANK-1	'person thanking' (all-matching nominatives)
	THANK-2	'recipient of thanks'
	THANK-3	'thing thanked for'

(10) I voted for a right-wing party.

| microroles: | VOTE-1 | 'voter' (all-matching nominatives) |
| | VOTE-2 | 'supported candidate' |

(11) I benefited from this change.

| microroles: | BENEFIT-1 | 'beneficiary' (all-matching nominatives) |
| | BENEFIT-2 | 'beneficial thing' |

(12) We're counting on a good result.

| microroles: | COUNT-ON-1 | 'person counting on sth' (all-matching nominatives) |
| | COUNT-ON-2 | 'thing counted on' |

(13) I told my brother about my problems.

microroles:	TELL-1	'speaker' (all-matching nominatives)
	TELL-2	'addressee'
	TELL-3	'topic'

(14) This belongs to my mother.

| microroles: | BELONG-1 | 'possessed thing' (all-matching nominatives) |
| | BELONG-2 | 'possessor' |

(15) We call this place 'the big meadow'.
microroles: CALL-1   'person naming sth' (all-matching nominatives)
                 CALL-2   'thing called a name'
                 CALL-3   'name attached to sth'
(16) I'm responsible for the performance.
microroles: RESPONSIBLE-1 'person bearing responsibility' (all-matching
                                     nominatives)
                 RESPONSIBLE-2 'scope of responsibility'
(17) He showed his children indifference.
microroles: SHOW-ATT-1   'person having a particular attitude' (all-matching
                                     nominatives)
                 SHOW-ATT-2   'person shown the attitude'
                 SHOW-ATT-3   'attitude shown'
(18) I'm looking at the young man.
microroles: LOOK-1   'observer' (all-matching nominatives)
                 LOOK-2   'thing looked at'
(19) I paid a fortune for this house.
microroles: PAY-1   'payer' (all-matching nominatives)
              PAY-2   'price' (all-matching accusatives)
              PAY-3   'thing paid for'
(20) I passed my brother his cup.
microroles: PASS-1   'person passing sth' (all-matching nominatives)
                 PASS-2   'recipient'
                 PASS-3   'thing passed' (all-matching accusatives)
(21) I showed the guests my town.
microroles: SHOW-1   'person showing' (all-matching nominatives)
                 SHOW-2   'person shown sth'
                 SHOW-3   'thing shown' (all-matching accusatives)
(22) I rely on my family.
microroles: RELY-1   'person relying on sb' (all-matching nominatives)
                 RELY-2   'person relied on'
(23) I need a new car.
microroles: NEED-1   'person having a need' (all-matching nominatives)
              NEED-2   'thing needed'
(24) I helped my father in his work.
microroles: HELP-1   'person helping' (all-matching nominatives)
                 HELP-2   'person helped'
                 HELP-3   'scope of help'

(25) The guest asked for a bottle of wine.

microroles: REQUEST-1 'person asking for sth' (all-matching nominatives)
             REQUEST-2 'thing asked for'

(26) I introduced my wife to a friend.

microroles: INTRODUCE-1 'person introducing sb else' (all-matching nominatives)
             INTRODUCE-2 'person introduced' (all-matching accusatives)
             INTRODUCE-3 'person acquainted with sb else'

(27) I'm preparing for work.

microroles: PREPARE-1 'person preparing' (all-matching nominatives)
             PREPARE-2 'thing prepared for'

(28) She accepted flowers from the man.

microroles: ACCEPT-1 'person accepting sth' (all-matching nominatives)
             ACCEPT-2 'thing accepted' (all-matching accusatives)
             ACCEPT-3 'donor'

(29) We reminded Tom about his promise.

microroles: REMIND-1 'person reminding' (all-matching nominatives)
             REMIND-2 'person reminded'
             REMIND-3 'thing reminded of'

(30) I asked the boy about the school.

microroles: ASK-1 'interrogator' (all-matching nominatives)
             ASK-2 'person asked'
             ASK-3 'topic of question'

(31) I explained the problem to the boss.

microroles: EXPLAIN-1 'person explaining' (all-matching nominatives)
             EXPLAIN-2 'issue explained' (all-matching accusatives)
             EXPLAIN-3 'recipient of explanation'

(32) I taught my son the Greek alphabet.

microroles: TEACH-1 'person teaching' (all-matching nominatives)
             TEACH-2 'learner'
             TEACH-3 'subject matter taught'

(33) I consider my neighbour a fool.

microroles: CONSIDER-AS-1 'person making a judgement' (all-matching nominatives)
             CONSIDER-AS-2 'person judged' (all-matching accusatives)
             CONSIDER-AS-3 'quality assigned by judgement'

(34) I believe in God.

microroles: BELIEVE-1 'person believing' (all-matching nominatives)
             BELIEVE-2 'thing believed in'

(35) We fought for freedom against the occupying army.

microroles: FIGHT-1   'person fighting' (all-matching nominatives)
                FIGHT-2   'thing fought for'
                FIGHT-3   'opponent in the fight'

(36) He looks like a beggar.

microroles: LOOK-LIKE-1   'person having an appearance' (all-matching nominatives)
                LOOK-LIKE-2   'quality suggested by appearance'

(37) I exchanged my old phone for a new model.

microroles: EXCHANGE-1   'person exchanging things' (all-matching nominatives)
                EXCHANGE-2   'thing exchanged for sth else' (all-matching accusatives)
                EXCHANGE-3   'thing acquired by exchange'

(38) It depends on his mood.

microroles: DEPEND-1   'dependent thing' (all-matching nominatives)
                DEPEND-2   'thing conditioning sth else'

(39) She assured her boyfriend of her love.

microroles: ASSURE-1   'person making an assurance' (all-matching nominatives)
                ASSURE-2   'person assured of sth'
                ASSURE-3   'thing assured of'

(40) My friend lacks money.

microroles: LACK-1   'person experiencing a lack'
                LACK-2   'thing lacking'

(41) Your stupid words led to an argument.

microroles: CAUSE-1   'causer' (all-matching nominatives)
                CAUSE-2   'caused effect'

(42) She found out about new job opportunities.

microroles: FIND-OUT-1   'person finding out sth' (all-matching nominatives)
                FIND-OUT-2   'information found out'

(43) He threatened the tourists with a knife.

microroles: THREATEN-1   'person making a threat' (all-matching nominatives)
                THREATEN-2   'person threatened'
                THREATEN-3   'instrument of threat'

(44) I'll inform the boss about our plan.

microroles: INFORM-1   'person giving information' (all-matching nominatives)

                INFORM-2   'recipient of information'

                INFORM-3   'topic of information'

(45) I gave back my sister her books.

microroles: GIVE-BACK-1 'person giving back' (all-matching nominatives)

                GIVE-BACK-2 'person given back their property'

                GIVE-BACK-3 'property given back' (all-matching accusatives)

(46) I remembered his birthday.

microroles: REMEMBER-1 'person remembering' (all-matching nominatives)

                REMEMBER-2 'thing remembered'

(47) We all desire a good solution.

microroles: DESIRE-1   'person having a desire' (all-matching nominatives)

                DESIRE-2   'thing desired'

(48) I offered the man a cup of tea.

microroles: OFFER-1 'person making an offer' (all-matching nominatives)

                OFFER-2 'potential recipient'

                OFFER-3 'thing offered' (all-matching accusatives)

(49) The committee awarded Tom a scholarship.

microroles: AWARD-1   'person awarding sth' (all-matching nominatives)

                AWARD-2   'recipient'

                AWARD-3   'thing awarded'

(50) I glanced at the cover of the book.

microroles: GLANCE-1   'person taking a glance' (all-matching nominatives)

                GLANCE-2   'thing glanced at'

(51) We were afraid for dad's health.

microroles: AFRAID-FOR-1 'person feeling fear' (all-matching nominatives)

                AFRAID-FOR-2 'thing endangered'

(52) My sister cares only about money.

microroles: CARE-ABOUT-1   'person caring about sth'

                CARE-ABOUT-2   'thing cared about'

(53) I'm telling you: watch out for this man.

microroles: WATCH-OUT-1 'person on guard' (all-matching nominatives)

                WATCH-OUT-2 'danger expected'

(54) The president answered all the questions.

microroles: ANSWER-1   'person answering' (all-matching nominatives)

                ANSWER-2   'utterance answered'

(55) Your offer is subject to assessment by our committee.
microroles: SUBJECT-TO-1   'thing subject to some rule' (all-matching
                                      nominatives)
                  SUBJECT-TO-2   'rule applied'
(56) He allowed the lovers a quick talk.
microroles: ALLOW-1   'person   giving   permission'   (all-matching
                            nominatives)
                  ALLOW-2   'recipient of permission'
                  ALLOW-3   'thing permitted'
(57) The company's director admitted to tax evasion.
microroles: ADMIT-TO-1 'person making a confession' (all-matching
                             nominatives)
                  ADMIT-TO-2 'confessed wrongdoing'
(58) They're searching for new employees.
microroles: SEARCH-1   'person searching' (all-matching nominatives)
                  SEARCH-2   'thing searched for'
(59) I mentioned your project to the professor.
microroles: MENTION-1 'person mentioning sth' (all-matching nominatives)
                  MENTION-2 'thing mentioned'
                  MENTION-3 'addressee'
(60) Which politician will Americans elect president in 2020?
microroles: ELECT-1   'person electing' (all-matching nominatives)
                  ELECT-2   'supported candidate' (all-matching accusatives)
                  ELECT-3   'position up for election'
(61) He uses the machine for various purposes.
microroles: USE-1   'person using sth.' (all-matching nominatives)
                  USE-2   'thing used' (all-matching accusatives)
                  USE-3   'purpose of use'
(62) I completely forgot about the party.
microroles: FORGET-1   'person forgetting' (all-matching nominatives)
                  FORGET-2   'thing forgotten'
(63) We're still considering your offer.
microroles: CONSIDER-1 'person considering sth' (all-matching nominatives)
                  CONSIDER-2 'topic of consideration'
(64) We finally decided on Spanish wine.
microroles: DECIDE-1   'person   making   a   decision'   (all-matching
                              nominatives)
                  DECIDE-2   'thing chosen by decision'

(65) I agree with Tom about this issue.
microroles: AGREE-1   'person agreeing' (all-matching nominatives)
                AGREE-2   'person agreed with'
                AGREE-3   'scope of agreement'
(66)) Suddenly the frog turned into a prince.
microroles: TRANSFORM-1   'thing undergoing transformation'
                              (all-matching nominatives)
                TRANSFORM-2   'new state formed by transformation'
(67) We all agreed on this price.
microroles: AGREE-ON-1 'person(s) agreeing' (all-matching nominatives)
                AGREE-ON-2 'thing chosen by agreement'
(68) He sold this bike to my brother for a small sum.
microroles: SELL-1   'seller' (all-matching nominatives)
                SELL-2   'product sold' (all-matching accusatives)
                SELL-3   'buyer'
                SELL-4   'price'
(69) Thick walls protect the house from the noise.
microroles: PROTECT-1   'person protecting' (all-matching nominatives)
                PROTECT-2   'thing protected' (all-matching accusatives)
                PROTECT-3   'danger'
(70) The teacher divided the students into three groups.
microroles: DIVIDE-1   'person dividing sth' (all-matching nominatives)
                DIVIDE-2   'thing divided' (all-matching accusatives)
                DIVIDE-3   'division created'
(71) The government deprived the poor family of their lands.
microroles: DEPRIVE-1   'person depriving sb' (all-matching nominatives)
                DEPRIVE-2   'person deprived of their property'
                DEPRIVE-3   'property taken away'
(72) I hid those documents from the inspectors.
microroles: HIDE-1   'person hiding sth' (all-matching nominatives)
                HIDE-2   'thing hidden' (all-matching accusatives)
                HIDE-3   'person disabled from finding sth'
(73) I sent my cousin a postcard.
microroles: SEND-1   'person sending' (all-matching nominatives)
                SEND-2   'recipient'
                SEND-3   'thing sent' (all-matching accusatives)
(74) He hit the attacker with a bottle.
microroles: HIT-1   'person hitting' (all-matching nominatives)
                HIT-2   'person hit' (all-matching accusatives)
                HIT-3   'weapon'

(75) He mistook my mum for my sister.

microroles: CONFUSE-1 'person   misjudging   sth'   (all-matching nominatives)

               CONFUSE-2 'thing misjudged' (all-matching accusatives)

               CONFUSE-3 'wrongly assigned identity'

Appendix 2

The following R code calculates the distance index for a pair of languages, as described in section 3.2. The text introduced by hash signs (until the end of the line of code) is ignored so that explanations can be intertwined with the code proper.

```
## The data must be provided as a csv file (example file
name used here: "cz-h.csv").
## Within the file markers of one language are stored
in one column.
## There are columns for two languages.
## The columns have headers (here: "Czech" and
"Hungarian").
## Neighbouring cells in one row contain markers of
the same microrole in the two languages.
## Another column with labels of microroles may be of
help, but it is not necessary for the computation.
## The folder containing the file must be made the
"working directory".
## We read the data into the programme.
czhdata <- read.csv("cz-h.csv")
c <- czhdata$Czech
h <- czhdata$Hungarian
czhtable <- table(c,h)
## We run the code that creates the function calcu-
lating the distance index.
distance.index <- function(dmat){
## This part is a function that takes a row or column
and returns the max (i.e. modal) value and the non-max
values.
## If two or more cells in a column or row have the
same max value, the formula takes one as max and all
other cells as non-max.
```

```
sumnosum   <-   function(x)c(max=max(x),nomax=sum(x)-
max(x))
## We loop the function over rows and columns.
maxr <- apply(dmat,1,sumnosum)
maxc <- apply(dmat,2,sumnosum)
## We compute the values for an evenly filled table
with the same marginal sums.
rs <- rowSums(dmat)
cs <- colSums(dmat)
xp <- rs %*% t(cs)/sum(rs)
## We do the same computation as for the actual table.
xpr <- apply(xp,1,sumnosum)
xpc <- apply(xp,2,sumnosum)
## This function does the summary comparison of max
and non-max values.
maxnomax <- function(x){
sum(x["nomax",])/sum(x["max",])
}
## We build and return results.
ireal <- c(index.mean = mean(c(indexr <- maxnomax(maxr),
indexc <- maxnomax(maxc))),
index.rows = indexr,
index.cols = indexc)
imax    <-    c(max.index.mean   =   mean(c(indexr
<- maxnomax(xpr),
indexc <- maxnomax(xpc))),
max.index.rows = indexr,
max.index.cols = indexc)
c(ireal,imax)
}
## We apply the function to our data.
distance.index(czhtable)
## The results include 6 values: "index.mean", "index.
rows", "index.cols", "max.index.mean", "max.index.
rows" and "max.index.cols".
```

References

Ágel, Vilmos/Fischer, Klaus: "Dependency Grammar and Valency Theory". In: Heine, Bernd/Narrog, Heiko (eds.): *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press 2009, pp. 223–256.

Aikhenvald, Alexandra Y.: "Grammars in Contact: A Cross-Linguistic Perspective". In: Aikhenvald, Alexandra Y./Dixon, R. M. W. (eds.): *Grammars in Contact. A Cross-Linguistic Typology*. Oxford: Oxford University Press 2006, pp. 1–66.

Bláha, Ondřej: *Jazyky střední Evropy*. Olomouc: Univerzita Palackého v Olomouci 2015.

Bühler, Karl: *Sprachtheorie: die Darstellungsfunktion der Sprache*. Frankfurt am Main/Berlin/Wien: Ullstein 1934/1978.

Campbell, Lyle: "Why is it so Hard to Define a Linguistic Area?" In: Hickey, Raymond (ed.): *The Cambridge Handbook of Areal Linguistics*. Cambridge: Cambridge University Press 2017, pp. 19–39.

Comrie, Bernard: "Slavonic Languages". In: Comrie, Bernard (ed.): *The World's Major Languages*. London/New York: Routledge 2009, pp. 269–273.

Gaszewski, Jerzy: *The Semantics of Selected Prepositions in Governed Prepositional Phrases. A Comparative Analysis of English-Polish Equivalences*. Ph. D. dissertation. University of Łódź 2012.

Gaszewski, Jerzy: "How to collect cross-linguistic data on valency online". In preparation a.

Gaszewski, Jerzy: "How to obtain a good sample of verbs". In preparation b.

Gaszewski, Jerzy: "Verb valency in Central Europe. A detailed look". In preparation c.

Hartmann, Iren/Haspelmath, Martin/Cysouw, Michael: "Identifying semantic role clusters and alignment types via microrole coexpression tendencies". In: *Studies in Language* 38, 3, 2014, pp. 463–484.

Haspelmath, Martin: "The European linguistic area: Standard Average European". In: Haspelmath, Martin/König, Ekkehard/Oesterreicher, Wulf/Raible, Wolfgang (eds.): *Language Typology and Language Universals*. Vol. II. Berlin/New York: de Gruyter 2001, pp. 1492–1510.

Haspelmath, Martin: "Comparative concepts and descriptive categories in cross-linguistic studies". In: *Language* 86, 3, 2010, pp. 663–687.

Haspelmath, Martin: "Transitivity prominence". In: Malchukov, Andrej/Comrie, Bernard (eds.): *Valency Classes in the World's Languages*. Vol. 1. Berlin/Boston: de Gruyter Mouton 2015, pp. 131–147.

Haspelmath, Martin/Hartmann, Iren: "Comparing verbal valency across languages". In: Malchukov, Andrej/Comrie, Bernard (eds.): *Valency Classes in the World's Languages*. Vol. 1. Berlin/Boston: de Gruyter Mouton 2015, pp. 41–71.

Herbst, Thomas/Heath, David/Roe, Ian F./Götz, Dieter: *A Valency Dictionary of English: A Corpus-Based Analysis of the Complementation Patterns of English Verbs, Nouns, and Adjectives*. Berlin/New York: de Gruyter Mouton 2004.

Hickey, Raymond: "Areas, Areal Features and Areality". In: Hickey, Raymond (ed.): *The Cambridge Handbook of Areal Linguistics*. Cambridge: Cambridge University Press 2017, pp. 1–15.

Karolak, Stanisław: "Walencja". In: Polański, Kazimierz (ed.): *Encyklopedia Językoznawstwa Ogólnego*. Wrocław/Warszawa/Kraków: Ossolineum 1999, pp. 629–630.

Koptjevskaja-Tamm, Maria: "Linguistic typology and language contact". In: Song, Jae Jung (ed.): *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press 2011, pp. 568–590.

Kurzová, Helena: "Mitteleuropa als Sprachareal". In: *Acta Universitatis Carolinae – Philologica* 5, *Germanistica Pragensia* XIII, 1996, pp. 57–73.

Newerkla, Stefan Michael: "Language Affinity in Central Europe – Some Thoughts on the Interrelations of German, Czech, Slovak and Magyar". In: *Opera slavica – Slavistické rozhledy* X, 4, 2000, pp. 1–16.

Newerkla, Stefan Michael: "Sprachliche Konvergenzprozesse in Mitteleuropa". In: Pospíšil, Ivo (ed.): *Crossroads of Cultures: Central Europe/Perekrestki kul'tury: Srednjaja Evropa/Křižovatky kultury: Střední Evropa*. Brno: Masarykova univerzita 2002, pp. 211–236.

Nichols, Johanna: *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press 1992.

Pilarský, Jiří: *Donausprachbund. Das arealistische Profil einer Sprachlandschaft*. Habilitation. University of Debrecen 2001.

Pusztay, János: "Közép-Európa nyelvi képe". In: Székely, Gábor/Jónás, Erzsébet (ed.): *Nyelvek és nyelvoktatás a Kárpát-medencében 1*. Nyíregyháza: Bessenyei György Könyvkiadó 1996, pp. 11–28.

Pusztay, János: "Central Europe as a landscape of convergence". In: *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 132, 2015, pp. 85–106.

Tesnière, Lucien: *Éléments de syntaxe structurale*. Paris: Klincksieck 1959/1976.

Thomas, George: "Exploring the parameters of a Central European Sprachbund". In: *Canadian Slavonic Papers / Revue Canadienne des Slavistes* 50, 1–2, 2008, pp. 123–153.

Jiří Januška

Central European Languages as a Complex Research Issue: Summarising and Broadening the Research Foci

Abstract: This paper examines the foci of research on Central European languages (German, Czech, Slovak, Polish, Hungarian, Croatian, and Slovene) and their convergence. It summarises the orientation of the research on this topic, illuminates the current shift in the orientation and outlines some propositions regarding its future orientation. Several main domains of the research are distinguished: the structural domain, the lexical domain, the phraseological domain, and the domain of language contact and language policy. In the section on the structural domain, a bibliographical overview on the topic of the idea of the Central European linguistic area (also called the Danube Sprachbund) is provided.

Keywords: linguistic convergence, Central Europe, theories of language contact, language change, Danube Sprachbund

1 Introduction[1]

The goal of this paper is to examine the foci of research on Central European languages (CEL)[2], i.e., of the research that aims to compare them and trace their convergence. On the one hand, the paper summarises the orientation of the research on this topic as it was conducted in the past decades. On the other hand, it illuminates the ongoing shift in the orientation and formulates some propositions concerning its future orientation.

The first question that arises is: *What are CEL*? One of the most common ways to search for the answer to this question, naturally, is to derive it from the answer to the question: *What is Central Europe (CE)*? The problem, however, is that there is no definite answer to this question. It is widely known that CE is

---

2   In this paper I will use these abbreviations: CE (Central Europe), CEL (Central European languages, the languages of Central Europe), CEA (the Central European linguistic area, the Danube Sprachbund).

a very ambiguous concept—even in its geographical sense, as Timothy Garton Ash's bon mot-like formulation shows:

> In an article published in 1954, the geographer Karl Sinnhuber examined sixteen definitions of Central Europe. The only part of Europe that none of them included was the Iberian peninsula. The only areas they all had in common were Austria, Bohemia, and Moravia. Tell me your Central Europe and I will tell you who you are. (Garton Ash 1999, p. 18)

Discussions of this kind concern the concept of CE not only in terms of geography, but also in terms of geopolitics, literature, various aspects of culture, etc. Three issues are then frequently considered, the individual definitions of which vary significantly: (1) what constitutes CE (i.e., borders, members, etc.);[3] (2) how should it be called;[4] (3) whether or not CE is an independent entity (i.e., whether its identity is constituted only by a geographical reference or whether its identity is constituted by a geographical reference *and simultaneously* by some other—geopolitical, literary, philosophical, etc.—properties that distinguish CE from other entities).[5]

It is not the aim of this paper to continue or to summarise these discussions.[6] The linguistic sense of CE seems to only partially be derived from these conceptions of CE (cf. section 3). To begin, let me state that the following languages, more or less, will be the primary subjects of this paper: German, Czech, Slovak, Polish, Hungarian, Croatian, and Slovene. These languages have a relatively long tradition of linguistic description that has brought a huge amount of findings on these languages, their comparison, and the exploration of their convergence. The research foci scrutinised by this paper mainly concern the

---

3   Among many others, in geopolitics there is also the concept of CE as a "crush zone between Russia and Germany" (O'Loughlin 2001, p. 614–616) or the "small nations" between Germany and Russia (Trencsényi 2017, p. 168).

4   Examples of English terminology used for the same (or almost the same territory) are: Central Europe, Central and Eastern Europe, East-Central Europe, Eastern Europe, Middle Europe, etc. It sometimes differs in different contexts (politics, culture, etc.) and it differs in different languages or countries. As a variation of Garton Ash's bon mot we could say: *Tell me how you call this territory and I will tell you who you are.*

5   As an example, the titles of some papers can be mentioned: "Does Central Europe exist?" (Garton Ash 1986), "Is there a Central-European identity in literature?" (Osers 2000), "Gibt es einen mitteleuropäischen Roman oder gibt es ihn nicht?" (Trávníček 2010), etc.

6   For such summaries see, e.g., Magocsi (2005), Gerhardt (2007), Trávníček (2009), Trencsényi (2017).

research which includes all (or almost all) of these languages together as a complex research issue.

As for the structure of the present paper, section 2 calls attention to another aspect of using the term *Central European*. Section 3 presents what is usually understood as CE in linguistics. Section 4 summarises the research issues that were in focus when comparing CEL and tracing their mutual influence (section 4.1.2 presents an extensive overview of the literature on the *Central European linguistic area*), and section 5 tries to illuminate a shift in the orientation of this research that is going on and to formulate some propositions concerning this future orientation.

2 'Laterality' of the Adjective *Central European*

This section explicitly takes into consideration another aspect of the concept of CE. General experience suggests that we should exercise a great deal of caution whenever dealing with an instance of the term *Central European*. We should always examine what the instance wants to refer to—not only what constitutes CE, i.e., what the entities referred to (regardless of whether we deal with countries or cultures or literatures or languages, etc.) are, but also what the indicated relations are among the referenced entities. Briefly speaking, e.g., two different senses of the adjective are used when statement A says that it deals with *Central European languages* and considers what is shared by all of them, and when statement B considers what is specific for each of them (but not necessarily shared by others). For both instances it is justified to use the adjective *Central European*, but each of them refers to a different relationship among the referenced entities.

I call this aspect *laterality* and I differentiate between four types of laterality:

1) *omnilateral*: concerning all members of the set;
2) *multilateral*: concerning three or more (but not all) members of the set;
3) *bilateral*: concerning two members of the set;
4) *unilateral*: concerning one member of the set.

In fact, the laterality can express different aspects. It seems reasonable to differentiate between *laterality of involvement* (i.e., how many languages are involved in the comparison), *laterality of connection* (i.e., how many languages are connected or share the same property), and the *laterality of actual contact* (i.e., how many languages are actually in contact with each other at once), etc.

For example, a paper that lists the loanwords of Slovak origin in each of the other CEL is omnilateral in regard to the laterality of involvement, but it is bilateral in the laterality of connection (it concerns only bilateral relations Slovak

> Czech, Slovak > Hungarian, Slovak > German, etc.). A paper describing the vowel harmony in Hungarian can also be labelled as dealing with the topic of *Central European* languages, but in its unilateral sense. A paper analysing the phonological features that are shared by all of the *Central European* languages and that distinguish them from other languages is omnilateral in its involvement and also in its connection. Of course, the complete conceptual typology of such kind is much more complex. However, this basic differentiation suffices for the rest of the present paper.

3 Central Europe as a Linguistic Landscape

In linguistics, the term *Central Europe* seems to be established for two territories: a wider one and a narrower one. As for the wider linguistic concept of CE, three authors can be mentioned. Ureland (2010) treats CE from the viewpoint of "a new areal linguistic orientation" which he calls *comparative Eurolinguistics*. Within this orientation, he presents the Eurolinguistic subdivision of Europe (Ureland 2010, p. 480, Fig. 1):

- West (Insular Celtic, North Sea Germ., Gallo-Rom.)
- South-West (Basque, Ibero-Rom., Hebrew, Romani, Arabic, Berber)
- North (North Germanic, Baltic Finnic)
- **Centre (Continental Germ., Baltic Lgs, W. Slavic, Hungarian, Alpine Rom.)**
- South (Gallo-Rom., Italo-Rom., Latin, Maltese, Arabic)
- East (Yiddish, East Slavic, Volga Finnic, Volga Turkic, Mongol)
- South-East (Balkan-Rom., South Slavic, Albanian, Greek, Turkish, Romani, Hebrew, Arabic)

The Eurolinguistic Centre is defined as "the European states now members of the European Union, situated between the Baltic and the Adriatic Seas" (Ureland 2010, p. 475). It is also divided from the Eurolinguistic East by the Catholic and Orthodox/Islam isogloss (Ureland 2010, p. 478; referring to a map in Wallace 1990).

For Pusztay (2015, p. 21), CE is "the belt stretching from the Alpine-Adriatic area to the Baltic area"; he calls it the *Amber Road Region*.[7] This region "lies between the German-spread western and Byzantine-Russian eastern civilisations, representing the eastern border of the western cultural circle" (Pusztay 2015, p. 16). It borders with Scandinavia in the north, with the Balkans in the south-east, and

---

7   It seems that Pusztay (2015) does not distinguish between terms like *region*, *area*, or *territory* as separate, precisely defined terminological units.

with the Mediterranean cultures in the south. Into this "landscape of linguistic convergence", the author counts the following languages: Finnish, Estonian, Livonian, Lithuanian, Latvian, Czech, Slovak, Hungarian, Croatian, Slovenian, and German (as the dominant donor language of the region). From the point of view of areal linguistics, the Amber Road Region consists of two linguistic areas (Pusztay [2015] uses the term *linguistic union*): the Baltic linguistic union and the Danube linguistic union. He has a terminological proposition: "The duality of the German terminology (Mittel-Europa [sic!] and Zentral-Europa [sic!]) can be eliminated if Mittel-Europa refers to the whole belt and Zentral-Europa is only a sub-region of it, that of the common territories of the Carpathian Basin and the Alpine-Adriatic area" (Pusztay 2015, p. 16).

Lastly, from the linguistic point of view, Hinrichs (2017) divides Europe into the Pan-European/West European Area, the Southeast European Area/Balkans, the Central European Area, and the East European Area. Then he subdivides Central Europe in the following way:

> On the basis of migration-related, cultural, economic, and linguistic factors, which have caused specific convergences, the modern space of 'Central Europe' can be divided into three areas, without laying down their borders too rigidly. Northern Central Europe is given the name 'Circum-Baltic Area' (Maria Koptjevskaja) and encompasses languages and cultures around the Baltic Sea. Middle Central Europe comprises today's Poland, Czech Republic, Slovakia, Hungary, and Germany, and occasionally Slovenia and parts of Ukraine. In cultural studies, a specific cultural space of *East Central Europe* (Ostmitteleuropa) is favoured (e.g., at the university of Leipzig). Southern Central Europe is rather an off shoot, comprising Croatia, as well as parts of Bosnia as far as the Adriatic. (Hinrichs 2017, p. 336)

The narrower concept of CE as a linguistic entity, which is the actual topic of the present paper, is called *Zentral-Europa* by Pusztay (2015) and *Middle Central Europe* by Hinrichs (2017).[8] It more or less covers the territory of the former Habsburg Empire. In linguistics, it is known as the *Central European linguistic area* (another frequent name is *Danube Sprachbund*) and most often comprises the following languages: German, Czech, Slovak, Hungarian, Croatian, and Slovene, sometimes Polish. The notion of this linguistic area is presented in section 4.1.2. of this paper.

---

8   Bláha (2015), whose CE generally corresponds to this narrower concept of CE, divides it into two subareas: the Sudetes subarea (Germany, Poland, Czechia, Austria, Slovenia) and the Pannonian subarea (Slovakia, Hungary, Croatia, Serbia, Bosnia and Herzegovina).

There seems to be no direct linguistic counterpart of the geopolitical concept of Central (and Eastern) Europe as the territory between Germany and Russia (cf. footnote 2). Of course, German and Russian—Europe's two biggest languages by number of native speakers[9]—represent the historically dominant linguistic influences in the territory of CE. Therefore, in some sense, these two languages can serve as referential points in comparing the distribution of (linguistic) similarities in this territory. According to what has been said above, the space between the continuous German language and Russian language territories can be divided into four parts: the Circum-Baltic space, Middle Central Europe, the Balkans, and the East Slavic languages (i.e., Belarusian, Ukrainian). The borders of this territory in the north and south are the seas; an inner border that can play a role (although in a fuzzy way) is the line between Latin and Orthodox Christianity.

4 A Review and Summary of Research Foci

CEL have a relatively long tradition of linguistic description. If we look at the linguistic literature on CEL, we may have the impression that the amount of findings is inversely proportional to the laterality (number of languages involved), i.e., the less laterality the larger the amount of the literature is: the largest amount of findings was produced on a particular CEL (i.e., descriptions of German, Czech, etc.), the smallest amount on relating all of the CEL together. In this context, we will not deal with unilateral and bilateral findings, i.e., the majority of the findings on CEL in detail.[10] We will instead *concentrate on omnilateral and multilateral findings, involving German, Hungarian, and several Slavic languages.*

Among all the findings on CEL, we can identify several main domains of research: the structural domain (i.e., phonology, morphosyntax), the lexical domain, the phraseological domain, and the domain of language contact and language policy. Naturally, other issues were also explored, but the mentioned ones seem to be the most often discussed and, hence, the most important.[11]

---

9   Cf. Tesnière (1928, p. 465) for data from the 1920's, for present situation cf. data from
    Lewis (2009, p. 545–580).

10 It is useful to keep in mind that any bilateral contrastive comparison of areally close
    languages has the potential to become a part of areal linguistic research. Bilateral
    findings on CEL will be mentioned only sometimes and marginally in this paper.

11 Besides these domains, other minor topics have also been in the focus of research.
    One of them is the question of mutual intelligibility which naturally concerns mainly
    genetically related languages; hence, in CE it concerns only the Slavic languages (for

There is not enough space here to report on the relevant literature in detail (with an exception in section 4.1.2).

4.1 The Structural Domain

The most consistent omnilateral structural correlation of CEL is the idea of the *Central European linguistic area* (CEA). Because there is no comprehensive bibliographical overview of the literature on this topic, it is presented in section 4.1.2 with brief summaries. Since the notion of *linguistic area* (*Sprachbund*) is considered to be a core concept of the field of areal linguistics, some remarks on this linguistic discipline have been made before, in section 4.1.1. Other omnilateral structural comparisons of CEL are mentioned in section 4.1.3.

4.1.1 Areal Linguistics and the Notion of Linguistic Area

The notion of *Sprachbund* (later *linguistic area*)[12] was originally proposed by Trubeckoj as a complementary concept to the notion of *language family*. It was coined into the linguistic discourse 90 years ago at the First International Congress of Linguists (Trubetzkoy 1928). It became one of the central research topics for the linguists associated with the Prague Linguistic Circle (PLC) like Trubeckoj, Jakobson, etc., and a continuous tradition of using this notion was initiated.

However, during these 90 years, the notion of linguistic area was used and defined in many different ways (for a commented overview see Campbell 2006a, pp. 7–17). Not only has the notion itself undergone various changes, but also the whole research field dealing with language convergence. We should keep in mind that contact linguistics as we know it today has developed long after the notion of Sprachbund was introduced. To illuminate the relationship between areal linguistics and contact linguistics (and other connected disciplines), I consider the model presented by Muysken (e.g., 2010) as very instructive, cf. Tab. 1. It is based on distinguishing several research levels that differ with regard to the space they cover, the time depth, typical types of data sources used for research, and the corresponding scenarios. Generally, different linguistic disciplines are concerned with different levels.

---

intelligibility of texts, e.g., Sloboda/Brankačkec (2014), Hofmański (2014); for intelligibility of lexemes cf. footnote 32).

12 Campbell (2006a, p. 3): "The name 'linguistic area' in English comes from Velten's (1943) translation of *Sprachbund* (literally 'language union'), made widely known by Emeneau (1956)." Sometimes, the terms *union* or *league* are also used in English.

Tab. 1: Four levels of aggregation and time depth in studying language contact (Muysken 2010, p. 268, Table 13.1)

	Space	Time	Source	Disciplines	Scenarios
Person	Bilingual individual	0–50 years	Recordings, tests, experiments	Psycholinguistics	Brain connectivity
Micro	Bilingual community	20–200 years	Recordings, fieldwork observations	Sociolinguistics	Specific contact scenarios
Meso	Geographical region	Generally 200–1,000 years	Comparative data, historical sources	Historical linguistics	Global contact scenarios
Macro	Larger areas of the world	Deep time	Typological data	Areal typology	Vague or no contact scenarios

In this model, areal linguistics—and thus also the notion of linguistic area—is connected to the meso-level.[13] We shall also notice that the levels in this model correlate with the number of languages involved as well (or as we could say: with laterality). Of course, the research of CEL is conducted at all levels (cf. for example the contributions in the present volume), but the omnilateral correlation of CEL, which is the main issue of this paper, happens at the meso-level.

Another important difference in areal linguistics to be mentioned, is that introduced by Campbell (e.g., 1985; 2017), namely the difference between the *circumstantialist approach* and the *historicist approach*:

> Different approaches have been taken in attempts to establish linguistic areas, with different implications for the definition of the concept. The *circumstantialist* approach (see Campbell 1985) mostly just lists similarities found in the languages of a geographical area and allows the list of shared traits to suggest diffusion and to define the linguistic area accordingly. Here, concrete evidence that the shared traits are diffused is not required. Circumstantialist areal linguistics has been criticized for not eliminating chance, universals, and possibly undetected genetic relationships among the languages as other possible explanations for shared traits. The *historicist* approach (Campbell 1985) seeks concrete evidence that the shared traits of an area are diffused (borrowed). This approach is more rigorous, although the lack of clear historical information in many

---

13 For the difference between areal typology, as it is understood in Muysken's table, and areal linguistics, see Dahl (2001, p. 1456).

cases has often led scholars to rely on the less trustworthy circumstantialist approach. (Campbell 2017, p. 24)[14]

It is worth mentioning that in dealing with CEA, different linguists employed different approaches.

4.1.2 CEA (Central European/Danube Sprachbund/Linguistic Area)

According to my bibliographical research, it was the Slovak linguist Ľudovít Novák who shaped the idea of CEA (*mitteleuropäischer Sprachbund*) for the first time. At the very beginning of World War II, he published two papers concerning *Central European comparative linguistics* (*porovnávacia jazykoveda stredoeuropska*) as he calls it (Novák 1939/40a, 1939/40b).[15] In the first one, Novák states that we can talk about the CEA as constituted by Czech, Slovak (West-Slovak and Central-Slovak), Hungarian, and German based on these two shared features: phonological quantity of vowels and unphonological stress on the first syllable.[16] Novák belonged to the younger members of the PLC, hence it

---

14 We should mention another possibility here, that the two different approaches can also be two different phases of research: a circumstantialist analysis can be a preparatory phase for a historicist analysis.

15 Both of them were published together shortly after the first appearance (Novák 1940).

16 Interestingly, the notion of CEA is not used in the Slovak main text of the paper at all, but only in its German and Russian summary (see the passage from the German summary): "Zusammenfassend: Man kann von einer mitteleuropäischen vergleichenden Sprachwissenschaft sprechen, deren Aufgabe es sein wird, im angegebenen Sinne die konvergente Entwicklung der Sprachen Mitteleuropas eingehend weiterzustudieren und so dem heute vom synchronistischen Standpunkt schon unzweifelhaften *mitteleuropäischen Sprachbund* auch eine historische Grundlage zu geben, in welchen nach der phonologischen Quantität und nach unphonologischen Akzent, der auf der ersten Silbe des Wortes allgemein wurde, heute das ganze čechische Gebiet gehört, das Westslovakische und Mittelslovakische und das Madjarische, die sich an das benachbarte Deutsche anschliessen, und zu denen früher (ungefähr bis zum XV. Jahrhundert) auch das Polnische, Lachische und Ostslovakische gehörte" (Novák 1939/40a: p. 101–102). English translation: "All in all, we can conclude that there is a Central European comparative linguistics, the aim of which is to extensively research the convergence between Central European languages. In this sense it should provide the *Central European Sprachbund*, which is from the synchronistic perspective not questioned anymore, with historical foundations, according to which the phonologic quality and the non-phonologic lexical stress (in common located in the first syllable) determine the inclusion of the entire contemporary Czech language area, the West and Central Slovak as well as Hungarian areas, which relate to the neighbouring German,

was natural that he focused on establishing *Sprachbünde*[17], as it was a new topic initiated by the linguists of the PLC, and it was natural that he mainly considered phonological features.[18]

Novák's paper did not establish a continuous tradition of using the notion, unlike Trubeckoj who proposed the general notion of *Sprachbund* in 1928. In the 1940s, we can find several linguists elaborating on the convergence of CEL, but none of them used Novák's notion of *mitteleuropäischer Sprachbund*. The first of them was the German Germanist and Fenno-Ugrist Ernst Lewy (1942) who presented a geographical-typological grouping of European languages. Among the five areas (*Gebiet*) postulated by him is the Central area constituted by German and Hungarian on the bases of word-flection (Lewy 1942, § 160–184). Since Czech and Polish evolved in close proximity to Russian, they were counted as the Eastern periphery of the Central area (Lewy 1942, § 350–353).

The Hungarian linguist and literary scholar László Gáldi (1946) compiled a patchwork of numerous particular partial similarities and convergences at different levels of the language system in the languages of East-Central Europe. As the most important languages of this part of Europe he listed: Polish, Romanian, Hungarian, Serbo-Croatian, Czech, Belarusian, Bulgarian Yiddish, Slovak, Slovene, and Albanian (Gáldi 1946, p. 113). Understandably, it is the Hungarian language that stood at the centre of his interest.[19]

The German Germanist Henrik Becker (1948) used the term *Sprachbund* (even in the title of the book), but he used it in a different sense than is usual. According to him, the *Sprachbund* is a group of languages that share(d) common cultural space and as a consequence "express almost the same in almost the same way" (Becker 1948, p. 5; translated by J. J.). He does not deal with phonological or

---

and historical (approximately to 15th century) Polish, Lachian and East Slovak areas." (Translated by J. J.)

17 He also established the *Mediterranean Sprachbund* (Novák 1932).

18 Papers by Novák (1934) and Skalička (1935) on the phonology of some CEL can be seen as a prologue to the introduction of the notion of CEA. Novák mentions the CEA also in his late papers (e.g., Novák 1979; 1984).

19 Gáldi is probably the first one in this context who uses the *Danube* as the principal label (his paper has the title *A Dunatáj nyelvi alkata* which means 'The linguistic makeup of the Danube region'). The reason for choosing it was not (only) linguistic but also political in some way. The paper was published in the first volume of the book *A Dunatáj* ('The Danube region') edited by a journalist and an officer of the Hungarian ministry of foreign affairs, Elemér Radisics. The goal of this book was to provide the Hungarian post-war audience with various information on countries surrounding Hungary in order to build better relationships with them.

morphosyntactic features, but only with loanwords and similar idioms. Europe is such a *Sprachbund* and the "Danube countries" constitute one of its *Unterbünde* (*subareas*). He says that 11–17 languages belong to this *Unterbund* and it was Viennese German that played the central role among them.

What was reported until this point of the current subsection can be seen as rather isolated contributions. On the contrary, in the 1970s we can observe the first intensification of coherent interest in this topic. The first paper of this period stems from the Czech typologist and Fenno-Ugrist Vladimír Skalička (1968). His study examines the convergence of the languages of the Danube basin (*Sprachen des Donaubeckens*) from the perspective of his theory of holistic typology (based dominantly—but not exclusively—on the basic morphological architecture of a language). In his frequently cited paper the term *Donausprachbund* is used only in the title, not in the text of the study itself. In the study, he actually analyses the CE *Sprachbund* (Hungarian, Slovak, Czech, and perhaps Serbo-Croatian and German)[20] and the Balkan *Sprachbund* (Romanian, Bulgarian, Macedonian, Albanian, and perhaps Modern Greek) and tries to answer the question: "Sind diese Sprachen nicht mehr als finnougrisch, romanisch, slavisch, sondern als balkanisch oder mitteleuropäisch zu bezeichen?"[21] (Skalička 1968, p. 3) At the end, his answer is: no. From the holistic typological view, these languages changed only very little; the typological dominants of these languages are more shared across their genealogical families than across their geographical areas.

The Hungarian Slavist and Fenno-Ugrist Gyula Décsy (1973)[22] proposed a new areal classification of the languages of Europe ('new' means different from Lewy's one). He divided them into eight zones (*Bund*).[23] One of them is the Danube Zone (*Donau-Bund*)—Czech, Slovak, Hungarian, Slovene, and Serbo-Croatian. He lists 18 shared features of various character. Décsy uses the term *Donau-Bund* in a different sense than Skalička (1968), namely for those languages that Skalička subsumed under the CE *Sprachbund*. Since Décsy (1973), the term *Danube Sprachbund* is used in this narrower "Austro-Hungarian sense" (i.e., without the Balkan languages).

---

20 As the features of the CE *Sprachbund*, he lists: initial stress, presence of palatal and affricate consonants, preposed article, compounds, three-tense system.

21 "Are these languages no longer to be called Finno-Ugric, Romance, Slavic, but Balkan or Central European?" (Translated by J. J.)

22 The English version of this book is Décsy (2000).

23 An important fact is that in Décsy's (1973; 2000) approach each language can be a member of only one *Bund*.

This narrowed sense of the term is also used by the German linguist and cultural scientist Harald Haarmann (1976, pp. 97–105; 1977a, pp. 8–9; 1977b, p. 114) who delimited the Danube *Sprachbund* (*Donausprachbund*) as one of the areal types in Europe. According to him, this *Sprachbund* is defined by seven shared features: phonological quantity in vowels, phoneme opposition /h/: /x/, fixed stress (usually initial), terminal consonant devoicing, synthetic nominal inflection, three-tense system, and productive (verbal) prefixation.

In his comprehensive paper, the Hungarian linguist and classical philologist János Balázs (1983) deals with the questions of mutual influence and convergence of languages in CE; he states that it is justified to speak about the existence of the Danube *Sprachbund* (*dunai nyelvszövetség*). The largest part of Balázs's study elaborates on specific partial influences among the CEL. He often stresses the role of Latin. Generally, the Hungarian language and the influence of other CEL on it are of special interest.

The intensification of the coherent interest in the areal perspective on CEL also brought about criticism of the notion of the CEA. The Hungarian Fenno-Ugrist István Futaky and his colleagues (Futaky et al. 1978) summarise their comments on 10 characteristic features of CEA (as specified by Skalička, Décsy, and Haarmann in their papers cited above). They argue that in CEL none of these features has developed under the influence of their neighbouring languages, but that these parallels are rather the result of the inner development in each particular CEL. Furthermore, some of the features are not unique, as they are also present in other languages that are genetically related to particular CEL.

The Hungarian Slavist, Hungarologist, Romanist and Africanist István Fodor (1983; 1984)[24] criticises Décsy's and Haarmann's argumentations as well. He comes to very similar final conclusions as Futaky et al. (1978). In the final part of his paper, he lists more than 30 salient structural features of Hungarian to demonstrate how different Hungarian is from other CEL. Fodor says that the idea of the CEA belongs "to the dustbin of the history of linguistics" (Fodor 1983, p. 66; 1984, p. 186).

In the second half of the 1990s we can observe the second intensification of interest in the topic, which has lasted until today. The Hungarian Fenno-Ugrist János Pusztay and his concept of the *Amber Road Region* (e.g., Pusztay 1996; 2003; 2015) were already mentioned above. According to him, this region (including the Baltic and the Danube unions) is characterised by the following linguistic features: the belt of composite languages, the belt of languages with

---

24 Fodor (1983) is a German version of Fodor (1984) which is written in Hungarian.

affix sequences, the belt of languages with preverbs, and unification with regard to case government. Additionally, he stresses a language sociological/political factor common for the CEL, namely the role of linguistic purism for language development and the ideological focus on the *mother tongue*.

The Czech Germanist Emil Skála (e.g., 1998)[25] was an enthusiastic proponent of CEA, however, he was interested in the similarities between only two languages: German and Czech. He also dealt with the history of Czech–German bilingualism and language contact.

The Czech classical philologist and comparative linguist Helena Kurzová (1996a; 1996b)[26] introduces—in an innovative way—some morphosyntactic features that should be considered in the areal linguistic research of CE: synthetic nominal inflection, synthetic comparison of adjectives, a simple system of tenses (i.e., no semantic opposition of two preterites), periphrastic future of an ingressive type (i.e., auxiliaries grammaticalised from verbs with ingressive meaning), formalised sentence structure with a finite verb and main actants in nominative-accusative form and/or position, limited use of the participle in gerundive function, agreement in an elaborated system of subordinate clauses, relative clauses with pronouns of interrogative origin, and productive preverbisation.

The Czech Germanist and Slavist Jiří Pilarský deals with the Danube *Sprachbund* in a series of papers (Pilarský 1995a; 1995b; 1996; 1997; 1998; 2000a; 2000b; 2001a), but first of all in his unpublished postdoctoral thesis (Pilarský 2001b). In this comprehensive study, the author presents an areal typological profile of the CE areal type which is constituted by these languages: German, Hungarian, Czech, Slovak, Serbo-Croatian, and Slovene. He examines the 12 structural features mentioned most frequently in the previous literature on this topic and he tries to establish their contrastiveness from the immediately neighbouring languages: Danish, Dutch, French, Italian, Albanian, Macedonian, Bulgarian, Romanian, Ukrainian, Polish, and Sorbian. He assumes a feature to be relevant for an areal type if it is not present in more than 50 % of the surrounding languages. Pilarský (2001b, p. 216) ends up with 8 structural features that define the CE areal type: stress on the first syllable, phonological opposition of quantity

---

25 Skála's papers on this issue (1991/1992; 1998; 1999; 2000; 2002; 2004) are de facto identical.

26 Kurzová dealt with the topic in the project "The Central Linguo-Cultural Area" (No. 831/91, Research Support Scheme of the Open Society Institute). The main result of this project was the manuscript "Contribution to European areal linguistics: defining the Central European area" (1995) which was never published. Unfortunately, in 2016 all attempts to find the manuscript in the archive of the Open Society Fund failed.

in vowels, absence of vowel reduction, productive verbal prefixation, synthetic nominal inflection, three-tense system, lexical convergence, and slight preference for prepositive attributes in NP.

The Austrian Bohemist and Slovakist Stefan Michael Newerkla (e.g., 2004), among others, stresses the pluricentric character of the German language. It should be considered that the convergence processes in CE involve different varieties of German as it is spoken in Austria, Bavaria, and elsewhere. Newerkla (2007a; 2007b) also proposes not to speak about one CEA, but about several CE contact areas (*Kontaktareale*).

The Canadian Slavist George Thomas (2008) defines the CE Sprachbund—or as he calls it in his paper the *Carpathian Sprachbund*—on the basis of five phonological and six morphosyntactic features: absence of phonemic pitch, fixed initial stress, phonemic quantity in vowels, no phonemic opposition of palatalisation, medial /l/, three-tense system, perfect as the sole preterite, periphrastic future with the verb 'become' as auxiliary, pluperfect formed with the double perfect, prepositive definite article with the demonstrative, and lastly prepositive indefinite article with the numeral 'one'. It involves the following languages: Czech, Slovak, Slovene, Kajkavian, Hungarian, Yiddish, and Bavarian-Austrian.

The book by Czech Slavist Ondřej Bláha (2015)[27] is a synthetic treatment of CEL. He deals with German, Czech, Upper and Lower Sorbian, Polish, Slovak, Hungarian, Slovene, and Croatian and either calls them "the languages of CE" (*jazyky střední Evropy*), "the languages of CEA" (*jazyky středoevropského areálu*) or the "CEA" (*středoevropský jazykový areál*). In separate parts of the book, the author discusses the cultural-historical aspects of CE, sociolinguistic characteristics of a particular CEL, gives a concise overview of the history and development of a particular CEL, and a concise overview of the language contact situation in CE, the lexicons of CEL—(mainly mutual) loanwords in CEL—, and almost 20 (more or less) common structural features of CEL (in word formation, syntax, inflection, and phonology). At the end of his book, the author concludes that his research did not approve of simplifying beliefs about linguistic convergence among CEL. The convergence took place by means of calquing, lexical borrowing and the semantic accommodation of a number of lexemes, also some word formation patterns and some types of complex sentences. On the contrary, such compact subsystems as inflection, phonology or simple sentence syntax have not undergone much convergence.

---

27 Also Bláha (2018).

The Russian Slavist Sergej Skorvid (2018) comments on the dominant role of German in CEA and on some of the features usually treated as characteristic of CEA. He also examines their diachronic development. Among others, he elaborates on an important idea concerning the time dimension of the convergence of CEL and its different stages—he distinguishes three of them: the earliest contact of tribal dialects, the Late Middle Ages, and the period of the Habsburg Empire).[28]

Of course, this is not an exhaustive list of authors and publications dealing with CEA. However, according to my bibliographical research these are the most important.[29] Tab. 2 and the following three paragraphs serve as a conclusion to this section of discussion.

The languages that are most often considered as constituting the CEA are:

- *core languages*: Czech, Slovak, Hungarian, and German (sometimes labelled as the dominant language of the area);
- *peripherical languages*: (Serbo-)Croatian, then Slovene;
- *marginal languages*: Polish and Yiddish.

The structural features that are considered constituting for the CEA are:

- *the most often*: fixed initial stress; phonological quantity in vowels; productive (verbal) prefixation; three-tense system; synthetic nominal inflection;
- *other structural features* (mentioned by at least two authors): rich possibilities of derivation and compounding in word formation (Skalička, Décsy, Haarmann, Pusztay, Bláha); convergence of verb valency (Pusztay, Newerkla, Bláha); periphrastic future of the ingressive type (Décsy, Kurzová, Thomas); synthetic comparison of adjectives and adverbs (Kurzová, Bláha); high number of morphonological alternations (Décsy, Bláha); wide use of periphrastic passive (Kurzová, Bláha); formalised sentence tending to the model found in Western European languages (Kurzová, Bláha, also Décsy); relative clauses with pronoun of interrogative origin (Kurzová, Bláha).

In summary, the reviewed papers define the horizon of the idea of a CEA as follows: They focus on searching for traits that are shared by all of the languages

---

28 Cf. Drinka's (2017) notion of Stratified Convergence Zone (section 5.1).

29 In my opinion, other publications worth mentioning are: Arany (1946/47), Nekula (1993; 1996; 2016), Hadrovics (1989), Honti (2000), Hamp (1989), Thomas (2010), Danylenko (2013).

in the linguistic area[30] and seek to answer the questions whether or not a CEA exists, which languages constitute it, and which features they share.

4.1.3 Other Structural Comparisons

Besides the idea of CEA, there has been no other consistent omnilateral research focus on the structural properties of CEL. This does not mean there are no such individual contributions. As examples of them, we can mention Hansen's (2005) brilliant study that deals with the modal systems of German, the Slavonic languages and Hungarian, Kiefer's (2010) paper on areal-typological aspects of aktionsart-formation in German, Hungarian, Slavic, Baltic, Romani and Yiddish, or Berger's (1995) contribution on the development of fixed stress in West Slavic languages with regard to German and Hungarian. Verb valency and the patterns of its similarities in CEL—an already aforementioned topic— are currently an issue of large-scale research projects (cf. the contributions by Gaszewski and Kim/Scharf/Šimko in the present volume). Among other multilateral comparisons of the structural properties of CEL, we can find a lot of works that concern Slavic languages or Slavic languages and German.

4.2 The Lexical Domain

The lexical domain is possibly the biggest domain of research on the CEL and their mutual influences. Amongst the long list of publications, the following prominently propose CE as a specific linguistic—in this case lexical—entity: Maćkiewicz (see 4.2.1), Unbegaun (see 4.2.2), and Půda (2010, p. 88: "Unser historisch lediglich lexikalisch durch Lehnwörter und Calques ausgelegter *Habsburger Bund* umfasst neben genetisch liierten auch geographisch benachbarte Sprachen über zentrale Donauregion hinaus."[31]). Using Matras' (2009) terminology, it concerns both matter replication (loanwords) and pattern replication (calques). Since it represents a giant amount of enumerative data, I only mention the references here.[32]

---

30 Sometimes this approach is extended with the centre–periphery structure (which, however, this does not change the basic logic: to find the core set of features that constitutes the prototype of the category).

31 "Our historical definition of the *Habsburger Bund* is based only on lexical factors, such as loanwords and calques, and hence includes not only genetically bound but also geographically close languages of the central Donau region." (Translated by J. J.)

32 Two marginal lexical subdomains could be mentioned that mostly apply to genetically related languages: the intelligibility of words or lexical distances (for Slavic languages

Tab. 2: The Central European linguistic area: terminology used by particular authors and the languages they regard as members of this linguistic area (ISO 639-3 codes stand for the language names; brackets mark the lower degree of membership of a language in CEA).

Author	Terminology	Languages involved						
		deu	hun	ces	slk	slv	(srp) hrv	others
Novák (**1939/40a**)	*mitteleuropäischer Sprachbund*	+	+	+	+	–	–	–
Lewy (**1942**)	*das zentrale Gebiet*	+	+	–	–	–	–	–
Gáldi (**1946**)	*Dunatáj, dunai táj, Keletközépeurópa*	+	+	+	+	+	+	pol, ron, bel, bul, yid, sqi
Skalička (**1968**)	*mitteleuropäischer Sprachbund; (Donausprachbund)*[a]	(+)	+	+	+	–	(+)	–
Décsy (**1973**)	*Donau-Bund*	–	+	+	+	+	+	–
Haarmann (**1976**)	*Donausprachbund*	+	+	+	+	–	(+)	–
Novák (**1984**)	*stredoeurópsky jazykový zväz*	(+)	+	+	+	–	–	(some rom dialects in Czechoslovakia)
Balázs (**1983**)	*dunai nyelvszövetség*	+	+	+	+	–	+	lat
Skála (**1998; 2000**)	*mitteleuropäischer Sprachbund, středoevropský jazykový areál*	+	(+)	+	(+)	(+)	(+)	[b]
Kurzová (**1996a; 1996b**)	*mitteleuropäisches Areal, Central European area*	+	+	+	+	(+)	(+)	(pol)
Pusztay (**1996; 2015**)	*Danube linguistic union, dunai szövetség; (Amber Road region)*[c]	+	+	+	+	+	+	–
Pilarský (**2001**)	*Zentraleuropäischer Arealtyp*	(+)	+	+	+	(+)	(+)	–
Thomas (**2008**)	*Carpathian Sprachbund (Central European Sprachbund)*	+	+	+	+	(+)	(+)	[d]

Tab. 2: (continued)

Author	Terminology	Languages involved						
		deu	hun	ces	slk	slv	(srp) hrv	others
Bláha (2015)	*středoevropský jazykový areál*	+	+	+	+	+	+	hsb, dsb, pol

[a] In Skalička's (1968) conception, the *Donausprachbund* consists of the Central European (*mitteleuropäischer*) *Sprachbund* and the Balkan (*balkanischer*) *Sprachbund*.

[b] Skála (1998; 2000) only works on Czech and German which he considers to be "the centre of the core of the Central European Sprachbund" (Skála 1998, p. 683; 2000, p. 84), but he does not mention what other languages constitute the *Sprachbund*. Since he quotes the shared features of CEA from Skalička (1968), other languages mentioned by Skalička (1968) are also given in the table (in brackets).

[c] In Pusztay's (1996; 2015) conception, the *Amber Road region* consists of the *Danube linguistic union* and the *Baltic linguistic union*.

[d] The core member of the *Sprachbund* is the Bavarian-Austrian dialect of German; the German standard belongs rather to the periphery of the *Sprachbund*.

4.2.1 Loanwords

The so-called *Europeisms*, i.e., loanwords shared across the languages of Europe, mainly of Latin and Greek origin, allow for an (almost) omnilateral comparison of loanwords in CEL. Maćkiewicz (e.g., 1992; 1996) adapts the notion of the linguistic area or linguistic league[33] to the lexical domain and proposes the notion of *lexical league* (*liga słownikowa, liga słownikowo-frazeologiczna*). In the frame of the European lexical league (*europejska liga słownikowa*) based on the aforementioned Europeisms, she tries to answer the question whether CEL constitute an autonomous area (Maćkiewicz 2004). She compares the Europeisms starting with the letter A in German, Polish, Slovak, Czech, Croatian, and Hungarian and finds more than 230 of them. From the analysis she concludes that there are some differences from Western European and Eastern European languages, but these differences are not striking. She also tries to find out how Polish (Maćkiewicz 1996; 2000) and Upper Sorbian (Maćkiewicz 2006) participate in the proposed European or Central European lexical league. Also Bláha (2015, pp. 75–95) deals with the *cultural Europeisms* in the languages of CE; he summarises the historical background of their borrowing and classifies them into several semantic domains.

The so-called *Carpathisms* represent a very specific lexical domain connected to the pastoral livelihood in the Carpathians. These lexemes can be found in

see Heeringa et al. 2013; Golubović/Gooskens 2015) and the so-called false friends (which are listed for particular language pairs).

33 In Polish, the term *liga językowa* is established as the primary term for a linguistic area.

Czech (Moravian), Slovak, Polish, Ukrainian, Hungarian, Romanian, Moldavian, Albanian, Macedonian, Serbian, and Croatian (Kloferová 2016). The principal opus of this research is *Obščekarpatskij dialektologičeskij atlas* (*The All-Carpathian Dialectological Atlas)* published in seven volumes 1989–2003 with preparatory and introductory volumes published in 1981 and 1987.

Another multilateral comparison of loanwords is the research on *Germanisms* in other CEL. With regard to the aforementioned pluricentric character of the German language, an important group of German words (and loanwords) is distinguished from German words (and loanwords) in general: the *Austriacisms*, i.e., the lexemes specific for the German spoken in Austria and not used in German spoken in Germany. Austriacisms in other CEL were compiled by Newerkla (e.g., 2007a; 2007b), several semantic groups of Austriacisms (religion, military, gastronomy, craftsman terminology) in Hungarian, Czech, and Slovak were described by Tölgyesi (e.g., 2010a; 2010b; 2011; 2012). Tóth (2011, pp. 58–60; 2012; 2013) compared the situation of shared Germanisms in today's Hungarian and Slovak. Germanisms in several Slavic languages were compared for example by Thomas (1997), Knoll (2008) or Kaczmarska/Kłos (2012). A huge amount of research findings on loanwords in CEL in general was published for particular bilateral relations between CEL.

4.2.2 Calques

Concerning multilateral lexical calquing in CEL, Ráduly elaborates on German calques in Polish and their equivalents in Russian, Ukrainian, and Belarusian (Ráduly 1999); on German calques in Hungarian, Croatian, and Polish (Ráduly 2002); and on German calques in Hungarian and Polish (Ráduly 1997). Nyomárkay (1980) deals with German calques in Hungarian and Polish; Jodas (2014) with German calques in Czech and their equivalents in Polish and Russian. Thomas (2003) studies German calques in four main Slavic languages of the former Habsburg Empire. Unbegaun (1932) explores calques in Slavic languages and comes to the conclusion that some Slavic languages prefer borrowing (Polish, Russian, Serbian, Bulgarian) while other Slavic languages prefer calquing (Croatian, Slovene, Czech, Sorbian)—he considers calquing to be the most characteristic property for CEL.[34] Again, the most research on calquing concerns particular languages pairs.

---

34 Unbegaun (1932, p. 48): "Aussi bien les Allemands, et à leur suite les Tchèques, les Sorabes, les Croates et les Slovènes, n'ont-ils pas été les seuls à utiliser le calque comme le moule normal à former des mots nouveaux. Il y a là un usage, sinon un mal, qui est aujourd'hui commun à toute l'Europe centrale et qui a atteint son plus

4.3 The Phraseological Domain

Phraseological parallels across CEL were shown by Newerkla (e.g., 2007a; 2007b), Tölgyesi (2013), or Fedosov (2002). Damborský (1977, pp. 82–89) compares Czech and Polish phrasemes; his study concludes that there is a tendency that Czech phraseology was rather influenced by German models while Polish phraseology was by French. Jodas (2014) uses a comparison with Russian that enables him to claim that the German influence is a question of degree on the scale Czech > Polish > Russian. There is a high number of other phraseological papers and dictionaries that focus on particular comparisons.

4.4 The Domain of Language Contact and Language Policy

The last among the most important and most often treated domain in literature on CEL is the topic of multilingualism and language policies and practice in CE. Descriptions of particular multilingual (bilingual) situations in CE are also present in many works mentioned above. Here, we only mention studies which primarily focus on the domain of language contact and language policy. Since this topic is most often treated for particular countries, i.e., states, it is natural that most languages are included in papers concerning the Habsburg Empire. For example, the volume named *Diglossia and Power: Language Policies and Practice in the 19th Century Habsburg Empire* (Rindler Schjerve 2003) contains a summarising exploration of the historical context, a methodological chapter about historical sociolinguistics and multilingualism, and then several case studies (e.g., the juridical system in Trieste [Czeitschner 2003], Ukrainian language in Galicia [Fellerer 2003], the education system in Plzeň [Newerkla 2003], etc.). Wolf (2015) deals with the multifaceted forms of translation and

---

grand développement chez les Magyars. On parle depuis longtemps de la communauté linguistique du monde balkanique, une communauté qui s'affirme par des traits généraux de vocabulaire et de syntaxe, voire de morphologie. Mais, si l'on s'attache quelque jour à déterminer une pareille communauté de l'Europe centrale, c'est le procédé du calque qui en sera l'indice le plus caractéristique." English translation: "The Germans, and after them the Czechs, the Sorbs, the Croats and the Slovenes, were not the only ones to use the calque as the normal way to form new words. There is a use, if not an evil, which is now common to whole Central Europe and has reached its greatest development among the Magyars. For a long time, we have been talking about the linguistic community of the Balkan world, a community that asserts itself with general vocabulary and syntax, and even morphology. But if we aim to determine such a community of Central Europe someday, it is the process of the calquing that will be its most characteristic attribute." (Translated by J. J.)

interpreting in the intensely pluricultural space of the Habsburg Empire between 1848 and 1918. The issue of language policies in the Empire or in one of its two parts (after 1867) was treated by Rindler Schjerve (2007; 2010), Marácz (2010), Fellerer (2011), etc.[35]

5   Broadening the Research Foci

In a nutshell, the review given in section 4 has revealed that concerning language structure, the principal foci of the reported research were, on the one hand, oriented towards structural features and, on the other, towards the lexicon. What is characteristic for the focus on structural features is the search for features which are shared by (more or less) all languages of the linguistic area and which are distinct for that linguistic area. The focus on the lexicon is characterised by very disparate and scattered results. Evidently, there is the need to broaden the foci and try to identify other similarity patterns of structural features and to aspire integration or better integrability of the various research on lexicon. This shift in focus (which is already partly ongoing) could lead to a more compact picture of the linguistic landscape of CE. This also holds true for the research on multilingualism in this part of Europe. It could make efforts to gather more historical information on multilingualism and to integrate them with what we know about other domains of the convergence of CEL. The following sections elaborate on these remarks.

5.1   Areal Linguistics without Linguistic Areas

The first thing I would like to pay attention to in this context is the necessity to *reconsider the notion of linguistic area (Sprachbund)*, since this notion was identified as the main research focus in the structural domain for CEL.

Generally, the notion of linguistic area can be useful to capture and express the simple idea that neighbouring, genetically (often) non-related languages show structural similarities in a couple of structural features. Such an idea can be useful in some general contexts. But yet, it is definitely not sufficient as a detailed statement of an areal linguistic analysis, similar to contemporary linguistic typology in which it is not sufficient to state that "Hungarian is an agglutinative language".

---

35 Particular bilateral situations of bilingualism, in particular countries of CE, are described in the second volume of Goebl et al.'s (1997) handbook.

If we consider the notion of linguistic area from the viewpoint of laterality (cf. section 2) and if we examine how much information it provides us with, we come to the following conclusion: It is an obvious fact that a group of languages shows mutual similarities of various types (and if these languages are in contact, then they show mutual convergences of various types): Some similarities are shared by two of the languages in focus (bilateral), other similarities are shared by more of them but not by all (multilateral), some similarities are shared by all of them—but also by other languages (omnilateral but not distinguishing), and some similarities can be shared by all of the languages in question and only by them (omnilateral and distinguishing). The idea of a linguistic area captures only the last type of information. If we exclusively focus on it, we lose the majority of information about the similarities and convergences among the languages in question. Hence, I propose to shift the focus from linguistic area-centred (*Sprachbund*-centred) analyses to a more detailed analyses of similarity patterns and convergence patterns in the structural domain of CEL.

This suggestion is completely consistent with the general recent development in areal linguistics. The notion of linguistic area has been criticised recently by researchers in the field and new or modified notions were proposed to replace it. I will quote a couple of examples of such criticisms and proposals.

Stolz (2006) sees an ontological problem with the notion of linguistic area or *Sprachbund*:

> *Sprachbünde* are not simply 'there'—they are constantly created anew by professionals. If the professional linguist feels that a certain number of shared features is necessary for the identification of a *Sprachbund*, this is largely a personal decision. (Stolz 2006, p. 36)

And he concludes:

> Thus, one should either strip the term of its unwelcome and much too suggestive connotations or abolish it for good (but it should be kept in the virtual museum of linguistic thought as an example of how difficulties and misunderstandings can be created via terminology). (Stolz 2006, p. 46)

Aikhenvald (2011) differentiates two levels within linguistic areal studies and she comments on the informativeness of the notion of macro-areas:

> Areal studies on a *macro-scale* are useful for a general view of what languages are like. But they communicate little about the precise history of the people involved, or the exact type of language interaction or the type of contact-induced change. They are also useful in orienting linguists towards the distribution of features and concentration of languages of particular type or with particular properties, or lack thereof. And they provide a basis for detailed descriptions and concentrating on smaller scale convergence zones.

Low-tier convergence zones within macro-areas allow us to establish the mechanisms by which matching structures develop, and also to see which features are resistant to being adjusted. This is where the knowledge of each others' languages, and contact between groups come into play. (Aikhenvald 2011, p. 18)

Campbell's (2006a, p. 1) opinion is that "it is the individual historical events of diffusion that count not the *post hoc* attempts to impose geographical order on varied conglomerations of these borrowings". He says:

A linguistic area, to the extent that it may have a legitimate existence at all, is merely the sum of borrowings in individual languages in contact situations. If we focus rather on understanding borrowings, those contingent historical events, the difficulty of determining what qualifies as a legitimate linguistic area ceases to be a problem. (Campbell 2006b, p. 459)

He examines various definitions of the notion of linguistic area (Campbell 2006a, pp. 7–17). His analysis also suggests that during the 90 years after its introduction by Trubeckoj, this notion has been used in so many different senses that it can hardly be regarded as unambiguous:

[…] not all areas are equal, or even similar, in their histories or composition. Things called 'linguistic areas' have included entities of widely divergent character and historical backgrounds, differing in social, cultural, political, geographical, attitudinal, historical and other factors; […]. (Campbell 2017, p. 22)

He concludes that "it would be more productive to investigate the facts of linguistic diffusion without the concern for defining linguistic areas" (Campbell 2017, p. 27) and he introduces a new notion (or rather a conceptual distinction). On one hand, he preserves the original notion, calling it *linguistic area sensu stricto* (LASS) "for a geographical region defined by shared diffused traits mostly contained within and shared across the languages of a clearly delimited geographical space" (Campbell 2017, p. 28). On the other hand, he coins the notion of *trait-sprawl area* (TSA):

The word 'sprawl' here reflects the fact that the individual traits can pattern in disordered ways, with some crisscrossing some languages while others crisscross other languages, with some extending in one direction, others in another direction, with some partially overlapping others in part of their distribution. (Campbell 2017, p. 28)

Some linguists working on the issue of the convergence of the Circum-Baltic languages also replaced the notion of *Sprachbund* with modified notions. Koptjevskaja-Tamm/Wälchli (2001) say:

Nau's main conclusion is that the CB [= Circum-Baltic] area is linguistically very complex, both synchronically and diachronically, with many layers of micro-and

macro-contacts and mutual influences superimposed on each other over a long period of time. Our guess is that intensive micro-contacts superimposed on each other some-times create an impression of an overall macro-contact among the languages in an area, which has not necessarily been there. We believe that the notion of Sprachbund tends to overemphasize the overall macro-contact, which might, of course, be justified in certain specific areas. For the CB area (and others comparable to it in the actual complexity of linguistic contacts), we suggest the term *Contact Superposition Zone*.
[…]
By abandoning the question of whether or not the CB languages constitute a Sprachbund, we can instead emphasize the most essential point in all areal linguistic studies, i.e. what kinds of areal convergence are found among the languages under consideration. (Koptjevskaja-Tamm/Wälchli 2001, p. 626)

In addition, they emphasise that there is no need to search for an omnilateral co-occurrence of traits in all of the languages we analyse:

Significantly, […] there are *no isoglosses covering all the CB languages*; moreover, the isoglosses pick up different subsets of the languages, in many cases also extending out-side of the CB area proper. Furthermore, there are only *few common innovations* in the area. It is rather the languages outside the CB-area, especially those to the south-west (SAE) that innovated structural properties. (Koptjevskaja-Tamm/Wälchli 2001, pp. 728, 732)

Drinka (2017) says the approach of Koptjevskaja-Tamm/Wälchli (2001) is extremely compatible with her own approach that stresses "the layered nature of that contact [i.e., the contact among languages in the Circum-Baltic—J. J.] across time and space" (Drinka 2017, p. 349). She says that the situation needs "a more fine-tuned depiction of complexity than the traditional image of a Sprachbund can provide" (Drinka 2017, p. 375) and proposes to replace it with the notion of *Stratified Convergence Zone* which is "a three-dimensional, chronologically strat-ified model" (Drinka 2017, p. 375).[36]

Hickey (2017) replaces the notion of linguistic area with *areality*:

The term 'linguistic area' can be a useful conceptual aid, and in the early days of research it helped to heighten scholars' awareness of shared structural features among not neces-sarily related languages in circumscribed geographical areas. However, the term came to dominate research (Campbell 2006), so that scholars often felt that a binary decision had to be made as to whether a given geographical area could be classified as a linguistic area or not. This concern has not always proved to be fruitful. What can be more significant is research into the forces and mechanisms which lead to languages in a given area coming to share features. This approach would highlight the scholarly concern with areality, that

---

36   In the context of CEL (or the CEA) it is Skorvid (2018) who stresses the stratification across time.

is, the areal concentration of linguistic features. How this concentration emerges and continues to develop is centre stage, not the attempt to attach the label 'linguistic area' in any given region. (Hickey 2017, pp. 1–2)

In this approach, areality is a matter of degree and it represents the areal concentration of linguistic features. Similar 'decomposition' of the notion of linguistic area in the context of CEL was sketched by Newerkla (2007a; 2007b):

> […] abychom spíše než o jednom středoevropském jazykovém areálu hovořili přesněji o několika středoevropských jazykových kontaktových areálech, rozdílně velkých, různě výrazných a vzájemně se někdy překrývajících. Stejně jako v dialektologickém bádání vymezují svazky izoglos určitá nářeční území, tak také svazky jazykověkontaktových jevů vzájemně oddělují jednotlivé kontaktové areály. (Newerkla 2007a, p. 275)[37]

According to Newerkla (2007a; 2007b), the linguistic area described by Kurzová (1996a; 1996b)—i.e., the more or less traditional concept of the CE linguistic area (cf. section 4.1.2)—is the central one among several others in CE. Even though Newerkla insists on the term *area* or *language contact area*, the core idea seems to be very similar or the same as the propositions above.

The mentioned opinions are only some examples for a view shared by various linguists, i.e., that the traditional notion of linguistic area should not be the focus of areal linguistic analysis anymore. The general conclusions are:

1) the traditional notion and the term linguistic area (*Sprachbund*) is insufficient and should be abandoned;
2) the focus should be shifted from exploring only the traits that are shared by all of the languages under consideration (or at least by the so-called core languages of the area) to more detailed exploration of the distribution of similarities and dissimilarities, convergences and not-convergences—and their patterns;
3) the new model should take account of stratification—in space, time, etc.

Besides the discussion about which one of the newly proposed notions would be the best to adopt for the case of CEL, we may want to have a neutral territorial notion expressing that we are interested in exploring the linguistic landscape in a

---

37 "[…] rather than to speak about one Central European linguistic area, we should speak more specifically about several Central European language contact areas that differ in size, distinctiveness and which sometimes overlap with each other. Similar to dialectological research, in which bundles of isoglosses delimit certain dialect territories, particular language contact areas are mutually separated by bundles of language contact phenomena." (Translated by J. J.)

particular territory. I propose to use the term *linguistic region*. Unlike a linguistic area or convergence zone that already states some specific kind or some specific (high) degree of common areality (if we use Hickey's [2017] terminology), a linguistic region would be a territory on which we try to explore the areality or the stratified convergence of the languages that are or were spoken in that territory—in other words: a territory on which we simply map the distribution of similarities and convergences (the patterns of clustering for different phenomena).[38]

A linguistic region is a geographic territory defined more or less arbitrarily (for example in consideration of some historical facts about borders or hypotheses about possible linguistic convergence in that territory) and it is open to the possibility to be broadened or narrowed. It is not a 'category' (such as language family or linguistic area), but rather a 'platform' for mapping the patterns of distribution.[39] Any linguistic region can be a subregion of a broader region at the same time.

In reference to Muysken's model (cf. section 4.1.1), the term linguistic region can be used for what he calls meso-level or macro-level. Within such a region there is the micro-level of bilingual communities and the person-level. Thus, we could for example talk about the *(Middle) Central European linguistic region* that is a territorial correlate of the CEA and which has its micro-level of actual bilingual communities (e.g., Croats in Burgenland, Hungarians in South Slovakia, etc.) and its person-level (which is a matter of psycholinguistics). This region is a subregion of the *Central European linguistic region* in a broader sense (cf. section 3).

5.2 Broadening the Array of Explored Language Phenomena

Another aspect to be considered is *the array of language phenomena we examine from the areal perspective*. It concerns what I called *domains* above. It became apparent that most of the effort in the complex research on CEL was devoted to structural features, lexical units (loanwords, calques), and phraseological units. But over the last decades, linguists have developed a number of new linguistic

---

38 It seems that, e.g., Koptjevskaja-Tamm/Wälchli (2001), Drinka (2017), Hickey (2017), and Campbell (2017) use the term *region* in this neutral territorial, geographical sense.

39 Thus, the traditional notion of linguistic area represents one of the possible similarity patterns that can be found in a linguistic region. The exploration should concern not only similarities, but also dissimilarities. (An interesting type of dissimilarity is mentioned by Bláha [to appear]: he calls it an "immune reaction", meaning preservation of some properties that are distinct from the properties of the dominating foreign language.)

'units'. They study new phenomena that are considered to function in the linguistic communication and to shape the language and its use. Some of them are: constructions (as Construction Grammar understands them with the whole range of their types, sizes, complexity, schematicity, etc.), collocations, prefabs (Erman/Warren 2000), cognitive construals, some reasonable variants of the so-called linguistic picture of the world—and we could continue to list them. Shouldn't the study of linguistic convergence reflect that?[40] For example, shouldn't we try to explore collocational calques, prefabs calques, construal calques etc. between/among CEL?[41]

5.3 Scale of Convergence and History of Multilingualism

The investigation of CEL as it was outlined in the previous paragraphs could also serve other purposes. One of them could be to contribute to the research on the so-called 'hierarchies of borrowability' or 'scales of adoptability' (Treffers-Daller 2010; Muysken 2010). Hickey (2017, p. 6) presents the scale of the degree to which particular linguistic levels are affected by convergence:

> *Levels of language most affected*
>> Vocabulary (loanwords, phrases)
>> Sounds (present in loanwords)
>> Speech habits (general pronunciation, suprasegmentals [stress, intonation])
>> Sentence structure, word-order
>> Grammar (morphology: inflection)
> *Levels of language least affected*

If we also include the 'new' phenomena (e.g., constructions of different types, collocations, construals etc.) into our research, we could maybe formulate hypotheses about *more refined* versions of such a scale of convergence.

Research on CEL explores not only the results and consequences of historical multilingualism in CE, but also the multilingualism itself—its forms,

---

40   Cf. a special issue of the *Journal of Language Contact* that is in preparation: "Usage-based contact linguistics: Effects of frequency and similarity in language contact". A question that suggests itself: Could we think of usage-based areal linguistics?

41   Such a broadening of focus could also concern the structural features: the research should focus not only on the presence of the features in a particular language, but also on their idiomatic usage. A good example is Schmiedtová/Sahonenko (2008), a psycholinguistic paper which shows that despite the fact that Czech shares the same aspectual structural feature with Russian (or generally with other Slavic languages), Czech speakers use different aspectual forms than the speakers of Russian when describing the same situation.

distribution, degrees, types, etc. An interesting innovation in this type of historical sociolinguistic research is gathering explicit data on multilingualism. As a synthesis, the two lines of research could be connected with each other: the elaborated scales of adoptability could be related to the real types and histories of multilingualism and be 'weighted' by them.[42] In other words, employing Muysken's (2010) scenario approach, which integrates social and structural factors in language change, the goal could be to investigate particular scenarios in the history and presence of CEL.[43]

5.4 Technological and Organisational Innovations

Another 'moral' of the research conducted until today is the capacity limitation: the capacity limitation of medium and the capacity limitation of individual researchers. The former means to reconsider the 'data storage' for the findings, i.e., their shift to the digital form. Today the printed paper form does not seem to be the appropriate medium for storing databases and other research results (i.e., the bibliographical, lexical, structural etc. data).[44] Also, some conceptual innovations could change the efficiency of the research: to build dictionaries of loanwords and calques as comparative databases or as some kind of 'lexical networks', etc.;[45] or—as a perhaps utopian idea—something like a "WikiWALS": a publicly accessible internet-based database of geographical, structural, constructional etc. properties of different languages that could be edited by linguists (and loaded with already published results of their research).[46]

---

42 An example of an obvious hypothesis is presented by Bláha (2018): the Czech administrative style is more converged with the German administrative style because of the role of German in the sphere of administration in the lands of the Bohemian crown until the formation of Czechoslovakia. Of course, more complicated findings could be reached.

43 For an example of research oriented in this way cf. the Special research programme *German in Austria. Variation – Contact – Perception* [*Deutsch in Österreich. Variation – Kontakt – Perzeption*] thoroughly described by Newerkla in the present volume.

44 Bibliographical databases seem to be an important requirement for integrating the research results which are often scattered in different journals and proceedings. See Newerkla and his team's project in this issue as proof of ingoing activity in this area.

45 Cf. the German Loanword Portal (*Lehnwortportal Deutsch;* http://lwp.ids-mannheim. de/) trying to collect various dictionaries of German loanwords in CEL in one database.

46 Cf. also other kinds of linguistic open data resources, e.g., Linguistic Linked Open Data (Web), Linguistic Data Consortium (Web), etc.

The latter capacity limitation concerns the fact that individual researchers have limitations restricting the possibilities of funding, number of languages involved in the research or number of languages in which the results of previous research on the topic were published, (non-compatible) methods used etc. A solution to this could be team work, the creation of research networks in which researchers could complementarily supplement each other and compensate out their individual limitations and biases.

6. Conclusion

The remarks made in this paper attempt to draw attention to (the need for) the process of integration and conceptual and technological innovations in research on CEL. The conceptual innovations correspond to the recent development in other branches of linguistics. Abandoning the notion of linguistic area can be seen as a similar development to what happened in linguistic typology, in which the model of holistic typology with its notion of language type was discarded and replaced by more fine-grained analyses, the contemporary model of which is called *distributional typology* (e.g., Bickel 2015). Involvement of 'new linguistic phenomena' correspond to the decomposition of the strict dichotomy of grammar–lexicon that took place in the last decades in linguistic theory. All of the propositions should provide us with a more informative, more efficient, and better organised model of the linguistic landscape in CE. The next step is to actualise these outlined propositions in particular studies.

References

Aikhenvald, Alexandra Y.: "Areal features and linguistic areas: Contact-induced change and geographical typology". In: Hieda, Osamu/König, Christa/Nakagawa, Hiroshi (eds.): *Geographical typology and linguistic areas: With special reference to Africa*. Amsterdam/Philadelphia: John Benjamins Publishing Company 2011, pp. 13–39.

Balázs, János: "Areális nyelvészeti kutatások története, módszerei és főbb eredményei". In: Balázs, János (ed.): *Areális nyelvészeti tanulmányok*. Budapest: Tankönyvkiadó/Frankin 1983, pp. 7–112.

Becker, Henrik: *Der Sprachbund*. Leipzig: Humboldt 1948.

Berger, Tilman: "Überlegungen zur Geschichte des festen Akzents im Westslavischen". In: Junghanns, Uwe (ed.): *Linguistische Beiträge zur Slawistik aus Deutschland und Österreich. II. JungslawistInnen-Treffen Leipzig 1993*. München/Berlin/Washington: Verlag Otto Sagner 1995, pp. 29–49.

Bickel, Balthasar: "Distributional typology: Statistical inquiries into the dynamics of linguistic diversity". In: Heine, Bernd/Narrog, Heiko (eds.): *The Oxford Handbook of Linguistic Analysis, 2nd edition*. Oxford: Oxford University Press 2015, pp. 901–923.

Bláha, Ondřej: *Jazyky střední Evropy*. Olomouc: Univerzita Palackého v Olomouci 2015.

Bláha, Ondřej: "Strukturní rysy jazyků střední Evropy". In: Pospíšil, Ivo (ed.): *Střední Evropa včera a dnes: Proměny koncepcí II*. Brno: Středoevropské centrum slovanských studií 2018, pp. 15–25.

Campbell, Lyle: "Areal linguistics and its implications for historical linguistic theory". In: Fisiak, Jacek (ed.): *Proceedings of the Sixth International Conference of Historical Linguistics*. Amsterdam/Poznań: John Benjamins Publishing Company/Adam Mickiewicz University Press 1985, pp. 25–56.

Campbell, Lyle: "Areal linguistics: A closer scrutiny". In: Matras, Yaron/McMahon, April/Vincent, Nigel (eds.): *Linguistic areas: Convergence in historical and typological perspective*. New York: Palgrave Macmilian 2006a, pp. 1–31.

Campbell, Lyle: "Areal Linguistics". In: Brown, Keith (ed.): *Encyclopedia of Language and Linguistics. 2nd Edition*. Oxford: Elsevier 2006b, pp. 454–460.

Campbell, Lyle: "Why is it so hard to define a linguistic area?" In: Raymond Hickey (ed.): *The handbook of areal linguistics*. Cambridge: Cambridge University Press 2017, pp. 19–39.

Czeitschner, Susanne: "Discourse, hegemony and polyglossia in the judicial system of Trieste in the 19th century". In: Rindler Schjerve, Rosita (ed.): *Diglossia and Power: Language Policies and Practice in the 19th Century Habsburg Empire*. Berlin/New York: Mouton de Gruyter 2003, pp. 69–106.

Damborský, Jiří: *Studia porównawcze nad słownictwem i frazeologią polską i czeską*. Warszawa: Państwowe Wydawnictwo Naukowe 1977.

Décsy, Gyula: *Die linguistische Struktur Europas*. Wiesbaden: Otto Harrassowitz 1973.

Décsy, Gyula: *The linguistic identity of Europe*. Bloomington: Eurolingua 2000.

Drinka, Bridget: *Language contact in Europe: The periphrastic perfect through history*. Cambridge: Cambridge University Press 2017.

Erman, Britt/Warren, Beatrice: "The idiom principle and the open choice principle". In: *Text: Interdisciplinary Journal for the Study of Discourse* 20(1), 2000, pp. 29–62.

Fedosov, Oleg: "Vengersko-neměcko-češskije frazeologičeskije paralleli (na fone aktualnych problem russkoj frazeologii)". In: *Rossica Olomucensia* 40, 2002, pp. 385–396.

Fellerer, Jan: "Discourse and hegemony: The case of the Ukrainian language in Galicia under Austrian rule (1772–1914)". In: Rindler Schjerve, Rosita (ed.): *Diglossia and Power: Language Policies and Practice in the 19th Century Habsburg Empire*. Berlin/New York: Mouton de Gruyter 2003, pp. 107–166.

Fellerer, Jan: "Multilingual states and empires in the history of Europe: The Habsburg Monarchy." In: Kortmann, Bernd/Van Der Auwera, Johan (eds.): *The languages and linguistics of Europe*. Berlin/Boston: De Gruyter Mouton 2011, pp. 713–728.

Fodor, István: "Verfügen die Sprachen des Donaubeckens über eine einheitliche Struktur? Zum Problem der Arealtypologie". In: *Finnisch-Ugrische Mitteilungen* (7), 1983, pp. 29–69.

Fodor, István: "Van-e a Dunatájnak egységes nyelvi alkata?" In: *Magyar nyelv* 80(1), 1984, pp. 36–47, 177–186.

Futaky, István et al.: "Der Donau-Sprachbund – gibt es ihn?" In: *Finnisch-Ugrische Mitteilungen* (2), 1978, pp. 181–192.

Gáldi, László: "A Dunatáj nyelvi alkata". In: Radisics, Elemér (ed.): *A Dunatáj: Történelmi, gazdasági és földrajzi adatok a Dunatáj államainak életéből*. Budapest: Gergely/Légrády Nyomda 1946, pp. 91–114.

Garton Ash, Timothy: "Does Central Europe exist?" In: *The New York Review of Books* 33, 9 October 1986, pp. 45–52.

Garton Ash, Timothy: "The Puzzle of Central Europe". In: *The New York Review of Books* 46, 18 March 1999, pp. 18–23.

Gerhardt, Sebastian: "Ostmitteleuropa – Zentraleuropa – Mitteleuropa: Annäherungen an Begriffe und ein Versuch zur Klärung ihrer Inhalte". In: *Rocznik Instytutu Europy Środkowo-Wschodniej* 5, 2007, pp. 33–43.

Goebl, Hans et al. (eds.): *Kontaktlinguistik: Ein internationales Handbuch zeitgenössischer Forschung*. Berlin/New York: Walter de Gruyter 1997.

Golubović, Jelena/Gooskens, Charlotte: "Mutual intelligibility between West and South Slavic languages". In: *Russian Linguistics* 39(3), 2015, pp. 351–373.

Haarmann, Harald: *Aspekte der Arealtypologie: die Problematik der europäischen Sprachbünde*. Tübingen: TBL Verlag Gunter Narr 1976.

Haarmann, Harald: "Arealtypologie der finnisch-ugrischen Sprachen – Prinzipien einer interdisziplinär orientierten Finnougristik". In: *Finnisch-Ugrische Mitteilungen* 1, 1977a, pp. 3–14.

Haarmann, Harald: *Prinzipielle Probleme des multilateralen Sprachvergleichs: Anmerkungen zur Methodik und Methodologie*. Tübingen: TBL Verlag Gunter Narr 1977b.

Hansen, Björn: "How to measure areal convergence: A case study of contact-induced grammaticalization in the German-Hungarian-Slavonic contact area". In: Hansen, Björn/Karlík, Petr (eds.): *Modality in Slavonic languages: New perspectives*. München: Verlag Otto Sagner 2005, pp. 239–257.

Heeringa, Wilbert et al.: "Lexical and orthographic distances between Germanic, Romance and Slavic languages and their relationship to geographic distance". In: Gooskens, Charlotte S./van Bezooijen, Renee (eds.): *Phonetics in Europe: Perception and production*. Frankfurt a. M.: Peter Lang 2013, pp. 99–137.

Hickey, Raymond: "Areas, areal features and areality". In: Hickey, Raymond (ed.): *The handbook of areal linguistics*. Cambridge: Cambridge University Press 2017, pp. 1–16.

Hinrichs, Uwe: "Linguistics". In: Mishkova, Diana/Trencsényi, Balázs (eds.): *European regions and boundaries: A conceptual history*. New York/Oxford: Berghahn 2017, pp. 322–349.

Hofmański, Wojciech: *Transfer ujemny a kompetencja językowa: Język polski w nauczaniu Słowian*. Praha: Karolinum 2014.

Januška, Jiří: *Porovnávání středoevropských jazyků: Za horizont strukturních rysů a lexikálních přejímek*. Doctoral thesis. Praha: Univerzita Karlova 2017.

Jodas, Josef: "Německé prvky ve slovní zásobě češtiny ve slavistickém kontextu". In: *Slavica Wratislaviensia* 159, 2014, pp. 169–177.

Kaczmarska, Elżbieta/Kłos, Zdzisław: "Grupy tematyczne zapożyczeń niemieckich w języku polskim, czeskim, słowackim i łużyckim. Problem analizy pożyczek pochodzenia niemieckiego w językach słowiańskich". In: *Zeszyty Łużyckie* 46, 2012, pp. 43–63.

Kiefer, Ferenc: "Areal-typological aspects of word-formation: The case of aktionsart-formation in German, Hungarian, Slavic, Baltic, Romani and Yiddish". In: Rainer, Franz et al. (eds.): *Variation and Change in Morphology: Selected papers from the 13th International Morphology Meeting, Vienna, February 2008*. Amsterdam/Philadelphia: John Benjamins Publishing Company 2010, pp. 129–148.

Kloferová, Stanislava: "Karpatismy v českém lexiku". In: Karlík Petr/Nekula, Marek/Pleskalová, Jana (eds.): *Nový encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny 2016, p. 785.

Knoll, Vladislav: "Výskyt německých dialektismů v západoslovanských jazycích". In: Příhoda Marek/Vaňková, Hana (eds.): *Slavistika v moderním světě*. Červený Kostelec/Praha: Pavel Mervart ve spolupráci s Filozofickou fakultou Univerzity Karlovy 2008, pp. 289–302.

Koptjevskaja-Tamm, Maria/Wälchli, Bernhard: "The Circum-Baltic languages: An areal-typological approach". In: Dahl Östen/Koptjevskaja-Tamm, Maria (eds.): *The Circum-Baltic languages: Typology and contact. Volume 2. Grammar and typology*. Amsterdam/Philadelphia: John Benjamins Publishing Company 2001, pp. 615–750.

Kurzová, Helena: "Mitteleuropa als Sprachareal". In: *Germanistica Pragensia* 13, 1996a, pp. 57–73.

Kurzová, Helena:"The future in Central European languages". In: *Listy filologické* 119, 1996b, pp. 4–18.

Lewis, M. Paul (ed.): *Ethnologue: Languages of the world*. Dallas: SIL International 2009.

Lewy, Ernst: *Der Bau der europäischen Sprachen*. Dublin: Hodges 1942.

University of Pennsylvania: *Linguistic Data Consortium*, retrieved 2.3.2018, from https://www.ldc.upenn.edu.

National University of Ireland Galway: *Linguistic Linked Open Data*, retrieved 2.3.2018, from http://linguistic-lod.org.

Maćkiewicz, Jolanta: "Wyrazy międzynarodowe a kształtowanie się europejskiej ligi słownikowej". In: Maćkiewicz, Jolanta/Siatkowski, Janusz (eds.): *Język a kultura. Tom 7. Kontakty języka polskiego z innymi językami na tle kontaktów kulturowych*. Wrocław: Wiedza o Kulturze 1992, 145–153.

Maćkiewicz, Jolanta: "Polszczyzna w europejskiej lidze językowej". In: Bobrownicka, Maria (ed.): *Symbioza kultur słowiańskich i niesłowiańskich*. Kraków: UNIVERSITAS 1996, pp. 197–202.

Maćkiewicz, Jolanta: "Czy Polska jest częścią Europy? Polszczyzna w europejskiej lidze językowej". In: Szpila, Grzegorz (ed.): *Język trzeciego tysiąclecia. Język a komunikacja 1*. Kraków: Tertium 2000, pp. 49–56.

Maćkiewicz, Jolanta: "Czy istnieje środkowoeuropejska wspólnota językowa? (na podstawie występowania europeizmów w językach Europy Środkowej)". In: Kątny, Andrzej (ed.): *Kontakty językowe w Europie Środkowej*. Olecko: Wszechnica Mazurska 2004, pp. 7–14.

Maćkiewicz, Jolanta: "Język górnołużycki wśród języków Europy Środkowej". In: *Lětopis* 53(1), 2006, pp. 107–114.

Magocsi, Paul Robert: "Eastern, East-Central, or Central Europe: Where is it and what is it?" In: *Kalendár = Almanac (National Slovak Society of the USA)* 113, 2005, pp. 128–140.

Marácz, László: "Multilingualism in the Hungarian Kingdom (1867–1918): Language policy and practice". In: Róka, Jolán (ed.): *Concepts & consequences of multilingualism in Europe*. Budapest: Századvég 2010, pp. 55–95.

Matras, Yaron: *Language contact*. Cambridge: Cambridge University Press 2009.

Muysken, Pieter: "Scenarios for Language Contact". In: Hickey, Raymond (ed.): *The handbook of language contact*. Malden (MA): Blackwell Publishing 2010, pp. 265–281.

Newerkla, Stefan Michael: "The seamy side of the Habsburgs' liberal language policy: Intended and factual reality of language use in Plzeň's educational system". In: Rindler Schjerve, Rosita (ed.): *Diglossia and Power: Language Policies and Practice in the 19th Century Habsburg Empire*. Berlin/New York: Mouton de Gruyter 2003, pp. 167–195.

Newerkla, Stefan Michael: "Rozmanitost jazyků a kultur v Rakousku-Uhersku a jejich konvergenční tendence". In: Pospíšil, Ivo/Moser, Michael (eds.): *Comparative Cultural Studies in Central Europe*. Brno: Masarykova univerzita 2004, pp. 11–42.

Newerkla, Stefan Michael: "Areály jazykového kontaktu ve střední Evropě a německo-český mikroareál ve východním Rakousku". In: *Slovo a slovesnost* 68, 2007a, pp. 271–286.

Newerkla, Stefan Michael: "Kontaktareale in Mitteleuropa". In: Kątny, Andrzej (ed.): *Słowiańsko-niesłowiańskie kontakty językowe*. Olecko: Wszechnica Mazurska 2007, 29–48.

Novák, Ľudovít: "De la phonologie historique romane: La quantité et l'accent". In: *Charisteria Guilelmo Mathesio quinquagenario a discipulis et Circuli linguistici Pragensis sodalibus oblata*. Praha: Pražský lingvistický kroužek (Cercle Linguistique de Prague) 1932, pp. 45–47.

Novák, Ľudovít: "Slovenčina a maďarčina v prvých storočiach ich dejín v strednej Europe". In: *II. międzynarodowy zjazd slawistów, Sekcja I. Księga referatów*. Warszawa: [s.n.] 1934, pp. 85–88.

Novák, Ľudovít: "Slovenské a podkarpatoruské nárečia vo svetle európskej fonologickej geografie: Synchronické i diachronické poznámky k porovnávacej jazykovede stredoeurópskej". In: *Linguistica Slovaca* 1–2, 1939/40a, pp. 85–105.

Novák, Ľudovít: "Neznáme nemecké vplyvy na západoslovanský, východoslovanský a maďarský konsonantizmus: Germanoslavistický príspevok k porovnávacej jazykovede stredoeurópskej (o zmenách (x) > h a ŕ > ř)". In: *Linguistica Slovaca* 1–2, 1939/40b, pp. 106–117.

Novák, Ľudovít: *Dve štúdie z porovnávacej jazykovedy stredoeuropskej*. Bratislava: Slovenská učená spoločnosť 1940.

Novák, Ľudovít: "Fonológia slovenčiny v kontexte inojazykových fonologických systémov". In: Mistrík, Jozef (ed.): *Studia Academica Slovaca. 8. Prednášky*

XV. letného seminára slovenského jazyka a kultúry. Bratislava: ALFA 1979, pp. 239–253.

Novák, Ľudovít: "Aký jazyk je slovenčina". In: Mistrík, Jozef (ed.): *Studia Academica Slovaca. 13. Prednášky XV. letného seminára slovenského jazyka a kultúry*. Bratislava: ALFA 1984, pp. 389–411.

Nyomárkay, István: "Német tükörfordítások a horvátban és a magyarban". In: *Magyar nyelv* 76, 1980, pp. 436–447.

O'Loughlin, John: "Geopolitical visions of Central Europe". In: Antonsich, Marco/Kolossov, Vladimir/Pagnini, Maria Paola (eds.): *Europe between geopolitics and political geography*. Rome: Societa Geografica Italiana 2001, pp. 607–628.

Osers, Ewald: "Is there a Central-European identity in literature?" In: Lord, Christopher (ed.): *Central Europe: core or periphery*. Copenhagen: Handelshjskolens forlag 2000, pp. 47–56.

Pilarský, Jiří: "A kvantitás-korreláció kérdése a dunai nyelvszövetségben". In: *Debreceni Szlavisztikai Füzetek* 1, 1995a, pp. 33–44.

Pilarský, Jiří: "Der phonetisch-phonologische Status der r-Laute in den Sprachen des Donausprachbundes". In: *Sprachtheorie und germanistische Linguistik* 1, 1995b, pp. 41–50.

Pilarský, Jiří: "Die Arealtypologie der orthographischen Systeme Europas und die Neuregelung der deutschen Rechtschreibung". In: *Sprachtheorie und germanistische Linguistik* 3, 1996, pp. 49–60.

Pilarský, Jiří: "Zum Status der Vokalquantität in den Sprachen des Zentraleuropäischen Areals". In: *Sprachtheorie und germanistische Linguistik* 6, 1997, pp. 27–39.

Pilarský, Jiří: "Zur arealtypologischen Ratio einer mehrdimensionalen Fortis-Opposition von gutturalen Frikativen innerhalb des Donausprachbundes". In: *Sprachtheorie und germanistische Linguistik* 8(2), 1998, pp. 177–203.

Pilarský, Jiří: "Zur Frage einer syntaktischen Konvergenz im Donau-Sprachbund. Struktur und Stellungstyp der Nominalphrase". In: *Sprachtheorie und germanistische Linguistik* 10(2), 2000a, pp. 167–198.

Pilarský, Jiří: "Zur Didaktik der deutschen r-Laute aus arealtypologischer Sicht". In: Viorel, Elena/Tóth-Nagy, Ildikó/Muncaciu-Codarcea, Emilia (eds.): *Beiträge zur Germanistik I*. Cluj-Napoca: Casa Cărții de Știință 2000b, pp. 145–162.

Pilarský, Jiří: "Auslautverhärtung als strukturtypisches Merkmal des Donausprachbundes?" In: *Sprachtheorie und germanistische Linguistik* 11(1), 2001a, 43–68.

Pilarský, Jiří: *Donausprachbund: Das arealistische Profil einer Sprachlandschaft*. Habilitation thesis. Debrecen: Debreceni Egyetem 2001b, retrieved at 2.3.2018, from: http://pilarsky.sweb.cz/dsb/Pilarsky-dsb.pdf.

Půda, Aleš: *Zur Theorie der Lehnprägung im deutsch-tschechischen Sprachkontakt: Eine historisch-vergleichende Untersuchung im innerslavischen und europäischen Kontext*. Frankfurt am Main: Peter Lang 2010.

Pusztay, János: "Közép-Európa nyelvi képe". In: Székely, Gábor/Jónás, Erzsébet (ed.): *Nyelvek és nyelvoktatás a Kárpát-medencében 1*. Nyíregyháza: Bessenyei György Könyvkiadó 1996, pp. 11–28.

Pusztay, János: *Közép-Európa:Nyelvi konvergenciatáj: Fejezetek a nyelvi egységesülés vizsgálatához*. Szombathely: Savaria University Press 2003.

Pusztay, János: *Convergent linguistic tendencies in the Bernsteinstrasse/Central-European linguistic area: Central Europe as a landscape of convergence*. Nitra: Constantine the Philosopher University in Nitra, Faculty of Central European Studies 2015.

Ráduly, Zsuzsanna: "Kalki niemieckie w języku polskim i węgierskim". In: *Studia Slavica Hungarica* 42, 1997, pp. 381–387.

Ráduly, Zsuzsanna: "Kalki niemieckie w języku polskim i ich odpowiedniki w języku rosyjskim, ukraińskim i białoruskim". In: *Studia Russica* 17, 1999, pp. 88–92.

Ráduly, Zsuzsanna: "Német tükörszavak a magyar, a horvát és a lengyel nyelvben". In: Jauk-Pinhak, Milka/Kiss, Csaba Gy. (eds.): *Croato-hungarica: Uz 900 godina hrvatsko mađarskih povijesnih veza*. Zagreb: Katedra za hungarologiju Filozofskog fakulteta Sveučilišta u Zagrebu/Matica hrvatska 2002, pp. 263–279.

Rindler Schjerve, Rosita (ed.): *Diglossia and Power: Language Policies and Practice in the 19th Century Habsburg Empire*. Berlin/New York: Mouton de Gruyter 2003.

Rindler Schjerve, Rosita: "Linguistic diversity in Habsburg Austria as a model for modern European language policy". In: Den Thije, Jan/Zeevaert, Ludger (eds.): *Receptive Multilingualism*. Amsterdam/Philadelphia: John Benjamins Publishing Company 2007, pp. 49–70.

Rindler Schjerve, Rosita: "What are the lessons to be learned from the language policy in the Western part of the 19th century Habsburg Empire?" In: Róka, Jolán (ed.): *Concepts & consequences of multilingualism in Europe*. Budapest: Századvég 2010, pp. 34–54.

Schmiedtová, Barbara/Sahonenko, Natascha: "Die Rolle des grammatischen Aspekts in Ereignis-Enkodierung: Ein Vergleich zwischen Tschechischen und Russischen Lernern des Deutschen". In: Gommes, Patrick/Walter,

Maik (eds.): *Fortgeschrittene Lernervarietäten: Korpuslinguistik und Zweitspracherwerbforschung*. Tübingen: Max Niemeyer Verlag 2008, pp. 45–71.

Skála, Emil: "Deutsch und Tschechisch im mitteleuropäischen Sprachbund". In: *Brücken* 1, 1991/92, pp. 173–179.

Skála, Emil: "Versuch einer Definition des mitteleuropäischen Sprachbundes". In: Ernst, Peter/Patocka, Franz (eds.): *Deutsche Sprache in Raum und Zeit: Festschrift für Peter Wiesinger zum 60. Geburtstag*. Wien: Praesens Verlag 1998, pp. 675–684.

Skála, Emil: "Der mitteleuropäische Sprachbund". In: Scharnhorst, Jürgen (ed.): *Sprachkultur und Sprachgeschichte*. Frankfurt a. Main: Peter Lang 1999, pp. 125–133.

Skála, Emil: "Středoevropský jazykový svaz". In: Kuklík, Jan (ed.): *Přednášky z XLIII. běhu Letní školy slovanských studií*. Praha: Filozofická fakulta Univerzity Karlovy 2000, pp. 77–85.

Skála, Emil: "Der mitteleuropäische Sprachbund". In: Scharnhorst, Jürgen (ed.): *Sprachkultur und Sprachgeschichte. 2., durchgesehene Auflage*. Frankfurt a. Main: Peter Lang 2002, pp. 125–133.

Skála, Emil: "Gibt es den mittelauropäischen Sprachbund?" In: *The Journal of Intercultural Studies* (Osaka) 30, 2004, pp. 22–29.

Skalička, Vladimír: "Zur mitteleuropäischen Phonologie". In: *Časopis pro moderní filologii* 21, 1935, pp. 151–154.

Skalička, Vladimír: "Zum Problem des Donausprachbundes". In: *Ural-Altaische Jahrbücher* 40(1–2), 1968, pp. 1–9.

Skorvid, Sergej: "Středoevropský jazykový svaz z hlediska jeho konstituování a konstituovanosti". In: Pospíšil, Ivo (ed.): *Střední Evropa včera a dnes: Proměny koncepcí II*. Brno: Středoevropské centrum slovanských studií 2018, pp. 177–187.

Sloboda, Marián/Brankačkec, Katja: "The mutual intelligibility of Slavic languages as a source of support for the revival of the Sorbian language". In: Fesenmeier, Ludwig/Heinemann, Sabine/Vicario, Federico (eds.): *Sprachminderheiten: Gestern, heute, morgen = Minoranze linguistiche: Ieri, oggi, domani*. Frankfurt am Main: Peter Lang 2014, pp. 25–44.

Stolz, Thomas: "All or nothing". In: Matras, Yaron/McMahon, April/Vincent, Nigel (eds.): *Linguistic areas: Convergence in historical and typological perspective*. New York: Palgrave Macmilian 2006, pp. 32–50.

Tesnière, Lucien: "Statistique des langues de l'Europe". In: Meillet, Antoine: *Les langues dans l'Europe nouvelle*. Paris: Payot 1928, pp. 291–484.

Thomas, George: "The Role of German Loanwords in the Slavic Languages of the Former Habsburg Empire". In: *Canadian Slavonic Papers/Revue Canadienne des Slavistes* 39(3/4), 1997, pp. 333–359.

Thomas, George: "Puristic attitudes to German phraseological and syntactic calques in the Slavic Languages of the former Habsburg Empire". In: *Canadian Slavonic Papers/Revue Canadienne des Slavistes* 45(1–2), 2003, pp. 201–225.

Thomas, George: "Exploring the parameters of a Central European Sprachbund". In: *Canadian Slavonic Papers/Revue Canadienne des Slavistes* 50(1–2), 2008, pp. 123–153.

Tóth, Sándor János: "Germanizmy v maďarsko-česko-slovenských súvislostiach". In: Pospíšil, Ivo/Zelenková, Anna (eds.): *Literární historiografie a česko-slovenské vztahy*. Brno: Tribun EU 2011, pp. 53–61.

Tóth, Sándor János: *Német–szlovák–magyar nyelvi összefüggések*. Szeged/Kassa/Brno: Gerhardus 2012.

Tóth, Sándor János: *Z funkčno-pragmatických aspektov kontaktovej a areálovej lingvistiky*. Ružomberok: Verbum 2013.

Tölgyesi, Tamás: "Náboženské výrazy řecko-latinského původu přejaté prostřednictvím němčiny do českého, slovenského a maďarského jazyka". In: Ondrejovič, Slavomír (ed.): *Varia XVII*. Ružomberok: Katolícka univerzita v Ružomberku/Filozofická fakulta/Slovenská jazykovedná spoločnosť pri SAV v Bratislave 2010a, pp. 470–474.

Tölgyesi, Tamás: "K německé řemeslnické terminologii v českém, slovenském a maďarském jazyce". In: Hladký, Juraj/Rendár, Ľubomír (eds.): *Varia XIX*. Trnava: Trnavská univerzita v Trnave 2010b, pp. 390–393.

Tölgyesi, Tamás: "K vojenskému názvosloví habsburské armády (1620–1918)". In: *Acta Universitatis Szegediensis. Dissertationes Slavicae: Sectio Linguistica* 29, 2011, pp. 71–79.

Tölgyesi, Tamás: "Német–cseh–szlovák–magyar nyelvi kontaktusok a közép-európai konyhaművészet tükrében". In: Kocsis, Mihály/Majoros, Henrietta (eds.): *Legendák, kódexek, források: Tanulmányok a 80 esztendős H. Tóth Imre tiszteletére*. Szeged: SZTE BTK Szláv Intézet 2012, pp. 331–338.

Tölgyesi, Tamás: "Jako ryba ve vodě. Wie ein Fisch im Wasser. Mint hal a vízben. Kontrastivní analýza českých, německých a maďarských frazémů o vodě". In: Zoltán, András/Fedosov, Oleg/Janurik, Szabolcs (eds.): *Voda v slavjanskoj frazeologii i paremiologii = A víz a szláv frazeológiában és paremiológiában = Water in Slavonic Phraseology and Paremiology II*. Budapest: Tinta Könyvkiadó 2013, pp. 633–638.

Trávníček, Jiří: "Zrození střední Evropy z ducha…" In: Trávníček, Jiří (ed.): *V kleštích dějin: Střední Evropa jako pojem a problém*. Brno: Host 2009, pp. 243–304.

Trávníček, Jiří: "Gibt es einen mitteleuropäischen Roman oder gibt es ihn nicht?" In: *Bayerisache Akademie der Schönen Künste, Jahrbuch* 23, 2010, pp. 65–75.

Treffers-Daller, Jeanine: "Borrowing". In: Fried, Mirjam/Östman, Jan-Ola/Verschueren, Jef (eds.): *Variation and change: Pragmatic perspectives*. Amsterdam/Philadelphia: John Benjamins Publishing Company 2010, pp. 17–35.

Trencsényi, Balázs: "Central Europe". In: Mishkova, Diana/Trencsényi, Balázs (eds.): *European regions and boundaries: a conceptual history*. New York/Oxford: Berghahn 2017, pp. 166–187.

Trubetzkoy, Nikolai: "Proposition 16". In: *Actes du premier congrès international de linguistes à La Haye, du 10–15 avril 1928*. Leiden: A. W. Sijthoff's uitgeversmaatschappij n.v. 1928, p. 18.

Unbegaun, Boris: "Le calque dans les langues slaves littéraires". In: *Revue des études slaves* 12(1–2), 1932, pp. 19–48.

Ureland, Sture: "Aspects of contact typology and linguistic areas in Central Europe between the Baltic and the Adriatic Seas". In: Hinrichs, Uwe (ed.): *Handbuch der Eurolinguistik*. Wiesbaden: PH 2010, pp. 475–502.

Wallace, William: *The transformation of Western Europe*. London: Pinter Publishers 1990.

Wolf, Michaela: *The Habsburg Monarchy's many-languaged soul: Translating and interpreting, 1848–1918*. Amsterdam/Philadelphia: John Benjamins Publishing Company 2015.

Agnes Kim

Prepositions in the Melting Pot: *High Risk of Infection*. Language Contact of German in Austria with Slavic Languages and Its Linguistic and Extra-Linguistic Description

Abstract: This paper investigates the (re-)construction of language myths in the linguistic and extra-linguistic discourse on language contact of German in Austria with Slavic languages. In a first step, it theoretically argues that individual language contact phenomena, such as various constructions with prepositions, may be items in the discourse archive of a superordinate language ideology. Through the analysis of the discourse on such language contact phenomena from the 19th to the early 21st century, this paper uncovers the underlying language myths that each tribe/nation has its own language (predominant in the 19th and early 20th century) and the one of the Habsburg monarchy as a linguistic melting pot (predominant after World War II). Additionally, the paper shows, how both myths are connected to an almost identical set of topoi, which are re-evaluated in the discourse after World War II. In a second step, this paper analyses individual language contact phenomena as language myths on their own, i.e., as reference points for common narratives. It observes the same processes of re-evaluation and proposes metalinguistic methods of historical contextualisation to exploit the linguistic and extra-linguistic discourse as the starting point for modern contact linguistic investigations and evaluations of the described contact phenomena.

Keywords: language contact, language myth, Habsburg monarchy, linguistic convergence, argument structure, PP-objects

1 Introduction[1]

Throughout at least the last 200 years, language has frequently been used to construct individual and group-specific, often "national" identities. In many cases, these identification processes happen *ex negativo*. Besides, language

contact and multilingualism play a particularly important role in the discursive construction of these identities. These aspects have gained special importance in linguistic ideologies, which (argumentatively) support language-based identities, as well as in language myths, which (narratively) underlie them.

This paper examines the role of language contact between German and the Slavic languages in Austria in the construction and representation of a specifically Austrian identity. The methodological approach of choice is a metalinguistic and discourse-analytic one as elaborated on in section 2. In the argumentation, we first examine the propagation of the positively assessed myth of *the Habsburg monarchy as a linguistic melting pot* after World War II. It is contrasted with the rather negative evaluation of language contact and multilingualism in linguistic and extra-linguistic publications during the Habsburg monarchy and Inter-War-Period (section 3). Finally, in section 4, the focus shifts towards single (alleged) contact phenomena, namely prepositional arguments. We treat them as language myths in their own right, i.e., as subjects to their own narratives (etymologies) and analyse the (re-)production of these narratives as well as their role in the linguistic and extra-linguistic discourse.

2 Theoretical and Methodological Approaches

This paper gives insight into the project parts of the Special research programme "German in Austria. Variation – Contact – Perception" (cf. Budin et. al. 2018), which are concerned with aspects of language contact of German in Austria with the Slavic languages (cf. Newerkla, this volume). Inter alia, project part 6 focuses on alleged contact phenomena that (are claimed to) have resulted from the intense contact of German in Austria (*Deutsch in Österreich*—subsequently: DiÖ) with various Slavic languages in general and especially Czech throughout the Habsburg monarchy until the end of World War I, and—to a certain degree—also the end of World War II. Its goal is to give a comprehensive overview and detailed analysis of contact induced Slavic influences on the varieties of DiÖ over time and, thus, also initiate the questioning of language myths and the correction of misperceptions and misjudgements with respect to these phenomena. A number of criteria allow for judgments on the plausibility of frequently cited contact explanations of single linguistic phenomena. In a next step they support the identification of *language myths* with regard to contact phenomena. These criteria group into (a) system linguistic, (b) sociolinguistic and (c) metalinguistic ones. In the following section, the criteria will be briefly described and essential terms will be explained.

2.1 Methodological Approaches to the Assessment of Contact Explanations

Generally, we are concerned with phenomena of language change in a multilingual society and thus, when reviewing alleged contact explanations, we need to consider both (a) language system internal factors as well as (b) language system external, i.e., sociolinguistic factors. As Hickey (2012, p. 403) puts it, "the linguistic course of change is connected with structural properties and developmental preferences which exist across languages and which ultimately have to do with language production and language processing". However, "the actuation, propagation, and conclusion of change is determined by social factors," which relate to the question of register, language, or—ultimately—feature choice and are thus considered to be even more complex in multilingual societies and communication settings.

(a) With regard to the assessment on the language system level, the application of methods from several linguistic subdisciplines depending on the nature of the phenomenon and the availability of contemporary or historical data is suggested. The choice of suitable methods requires a detailed description of the phenomenon with regard to its (alleged) diatopic and diastratic distribution and even a reassessment of the linguistic level. In section 4, this article gives insight into the thorough research conducted in order to facilitate in-depth studies of single or groups of contact phenomena. An exemplary study with a system linguistic focus on such an alleged contact phenomenon, namely the case variation of the verb *vergessen* 'to forget' in DiÖ, which employs contrastive and corpus linguistic methods, can be found in this book (cf. Kim/Scharf/Šimko, this volume).[2]

(b) Additionally, the sociolinguistic circumstances of language contact require consideration. Domain-specific[3] and multi-source approaches help to

---

2 Of course, not all alleged contact phenomena in DiÖ can be analysed in detail within the project. Therefore, its main aim is to provide a comprehensive overview of the phenomena, which includes a detailed system linguistic description, a summary on its (alleged) distribution and suggestions about possible methodological approaches for detailed analyses. This information will be provided online within the Information system on (historical) Multilingualism in Austria (*Informationssystem zur [historischen] Mehrsprachigkeit in Österreich* – MiÖ, cf. Kim et al. [Web]) at the Collaborative Online Research Platform of the SFB "German in Austria: Variation – Contact – Perception".

3 A domain-specific approach to linguistic practices in multilingual societies presupposes that multilingualism is functionally organised, i.e., *diglossic* (cf. Ferguson 1959). It is thus possible to abstract *domains of language use*, such as the school or the family from

describe the interaction of language policies and language use and to recon-struct local and regional *language contact scenarios* (cf., e.g., Newerkla 2003; Kim/Newerkla 2018; Kim 2019). According to Muysken (2010, p. 267), a language contact scenario is "the organized fashion in which multilingual speakers, in certain social settings, deal with the various languages in their repertoire." Knowledge of these usage-based principles enables predictions regarding the kind of contact phenomena that are most likely to occur. Hence, a combination of the domain-specific and the scenario approach satisfies the need to cope with the interconnectedness of language system internal and language system external factors in linguistic change.

(c) Eventually, besides these system linguistic and sociolinguistic factors a third, metalinguistic one needs to be taken into account when it comes to the as-sessment of the plausibility of contact explanations: Mailhammer (2013, p. 11) describes the so-called *Internal Development Bias*, i.e., a practice in historical linguistics to prefer language internal explanations over language contact explanations for lexical items. This is especially true for structur-alist approaches, such as the one employed by the Viennese dialectolog-ical school, which conducted the vast majority of scientific investigations into German in Austria throughout the 20th century. Even if sociolinguistic circumstances of language change were taken into account, language con-tact explanations were frequently either marginalised (e.g., cf. Seidelmann 1971) or openly rejected (e.g., cf. Ernst 2008). Hence, a thorough historical contextualisation of the publications, which constitute and/or contribute to the dominant discourse on single contact phenomena, is indispensable. A description of the development and interaction of such discourses will probably enable the tracing of the shift in the dominant linguistic ideology that took place in Austria after World War II.

2.2 Theoretical Approach: Language Myths and Linguistic Ideologies

This article draws upon the theoretical concept of language myths as propagated by Watts (2012). According to him, *language myths* are "communally shared stories" (Watts 2012, p. 589) which provide "a narrative cognitive embedding of a belief, or sets of beliefs, about some aspect of a socio-cultural group" and "help to set up a foundation for performing acts of identity in emergent social practice"

---

specific social settings, such as a specific school or family (cf. Rindler Schjerve 1996, p. 797; Rindler Schjerve [eds.] 2003).

(Watts 2012, pp. 600–602). The *narrative structure* of these myths comprises a "restricted set of conceptual metaphors generating 'true' statements." Such language myths, however, structure the general as well as scientific discourse on language. Thus, they instantiate *linguistic ideologies*, which gain their special explanatory power if perpetuated by professional linguists or linguistic genres such as dictionaries or grammars. As products of the dominant discourse they produce and reproduce knowledge and, simultaneously, the "'laws of what can be said' about language", i.e., the *discourse archive* (Watts 2012, pp. 589, 600–602).

In this light, single linguistic (alleged) contact phenomena in DiÖ are items of such a discourse archive: As will be shown in section 3, they draw upon the independent linguistic history of DiÖ within the multilingual Habsburg monarchy, in which language contact specifically shaped German and hence made it a distinct variety of German. This ideology of a *distinct Austrian German* and the underlying ideology of a *distinct Austrian nation* developed in the aftermath of World War II in order to renounce the *großdeutsch*-oriented ideology of the *cultural and linguistic German nation*. Until then, the latter ideology had been dominant in the construction of a national identity in what is nowadays Austria. A dissociation was necessary, because it had culminated into Austria's denied collective complicity in the Nazis' crimes (cf. de Cillia/Wodak 2006; Glauninger 2015). The myth of the *Habsburg monarchy as a linguistic melting pot* supports and thus— amongst others—underlies this ideology. Section 3 describes it as propagated in the 19th, 20th and 21st century linguistic and extra-linguistic discourse.

However, each single linguistic (alleged) contact phenomenon may also be treated in its own right, i.e., as the centre of the specific scientific and extra-scientific discourse, which constructs it as a linguistic phenomenon specific for DiÖ and influenced by historical language contact. The accompanying narratives on the single phenomena do not originate from the second half of the 20th century but have rather been continuously repeated from the 19th century onwards and then aggregated to establish the above-mentioned ideology of a *distinct Austrian German*. Section 4 of this article establishes an approach to specific contact phenomena as the subject of narratives, i.e., as language myths, and exemplarily elaborates on the (re-)production of such stories in the example of *prep*-arguments.

3 The Habsburg Monarchy as a Linguistic Melting Pot

Section 3 traces the myth of *the Habsburg monarchy as a linguistic melting pot* and its exploitation in the construction of a specific Austrian identity backwards. Firstly, it offers a description of the *status quo* in the late 20th and early 21st

century. Consequently, it describes the development throughout the second half of the 20th century as reflected in dictionaries of DiÖ. Finally, it focuses on the linguistic and journalistic discourse in the late 19th and early 20th century.

3.1 The Linguistic Melting Pot in Contemporary Linguistic Ideologies in Austria

The very first constitution of the First Austrian Republic (BGBl. Nr. 1/1920, Art. 8) declared German the national language of the young state. The linguistic minorities had to be granted special linguistic rights according to the Treaty of Saint-Germain-en-Layes. As late as 2000, the parliament agreed on a complementary paragraph, which declared the protection and support of the autochthonous ethnic groups a state objective (Germ. *Staatszielbestimmung*). It reads:

> Die Republik (Bund, Länder und Gemeinden) bekennt sich zu ihrer gewachsenen sprachlichen und kulturellen Vielfalt, die in den autochthonen Volksgruppen zum Ausdruck kommt. Sprache und Kultur, Bestand und Erhaltung dieser Volksgruppen sind zu achten, zu sichern und zu fördern. (BGBl. I Nr. 68/2000)[4]

The notion of "linguistic and cultural multiplicity having grown" conceptualises multilingualism in Austria as essentially historical: Austria is a monolingual German country that cannot deny its historical multilingualism, which is still present in the society today and thus needs to be embraced and incorporated into the Austrian identity. However, that incorporation is restricted to the recognised ethnic groups, which "so to speak shared our [the Austrians'] fate in the last centuries", as Harald Ofner,[5] a parliamentarian representing the FPÖ[6], put it in the parliamentary debate before the vote on the respective state objective.

---

4   The Republic (Federation, provinces and municipalities) subscribe to its linguistic and cultural multiplicity having grown, expressed in the autochthonous ethnic groups. Language and culture, existence and preservation of these ethnic groups are to be respected, safeguarded and to be supported. (Official translation according to Federal Constitutional Law [Web]).

5   Harald Ofner (*25.10.1932 Wien), cf. Republik Österreich, Parlament (Web). Ofner had served as the minister of justice during the coalition of his party with the Social Democratic Party of Austria (*Sozialdemokratische Partei Österreichs*, SPÖ) from 1983 to 1987.

6   The Freedom Party of Austria (*Freiheitliche Partei Österreichs*, FPÖ) is considered a right wing populist and national-conservative political party. From 2000 to 2005, the FPÖ formed a coalition with the Austria's Peoples Party (*Österreichische Volkspartei*, ÖVP).

This example excellently illustrates how the historical multilingualism that shaped the Habsburg monarchy as a whole, and thus the Austrian crown lands too, serves as an element for the construction of a distinct Austrian identity. Similarly, it reoccurs in linguistic literature as a *locus communis*: Pohl (1997, p. 1811) insists that the Austrian scientific community agrees on the fact that the minorities, i.e., autochthonous ethnic groups are a constitutive element of modern Austria and justifies the following statement:

> Denn Österreich hat auf seinem Weg ins 20. Jahrhundert mehrere schwierige Stationen durchlaufen, deren entscheidende jene der Habsburger-Monarchie war, nach deren Untergang sich das verbleibende „Deutschösterreich" 1918 als unglücklicher Kleinstaat wiedersah, der erst nach 1945 zu sich selbst gefunden und ein neues (politisch nicht-deutsches) Selbstbewußtsein entwickelt hat, in dem die Minderheiten als ein bereicherndes Element und wertvolles Erbe aus früherer Zeit ihren Platz haben. (Pohl 1997, p. 1811)[7]

Again, the autochthonous ethnic groups are highlighted as an enriching element and valuable heritage that connects modern Austrian identity to its history as the centre of the Habsburg monarchy. With regard to linguistic phenomena, Pohl (1997, p. 1808) claims that in the standard register of DiÖ, language contact only left its traces at the lexical level. Nonetheless, the myth of the *Habsburg monarchy as a melting pot* plays a constitutive role for the Austrian linguistic identity.

Similarly, language contact and historical multilingualism are commented on in Muhr/Schrodt/Wiesinger (1995), the proceedings of a conference on various aspects of Austrian German. The book comprises three sections, one of which naturally deals with aspects of language contact and especially language contact with Slavic languages. The paper, which focuses on the linguistic contact of Czech and German (cf. Spačilová 1995), opens with the following remark:

> Das langjährige Zusammenleben der Tschechen mit den Österreichern in der multinationalen habsburgischen Monarchie beeinflußte ohne Zweifel in vielen Hinsichten die Gewohnheiten beider Völker. (Spačilová 1995, p. 327)[8]

---

7   Because Austria passed through a number of difficult stages on its way into the 20th century, the most decisive of which was the Habsburg monarchy. After its downfall in 1918, the remaining "Deutschösterreich" reappeared as an unfortunate small state, which did not find itself until after 1945, when it developed a new (politically non-German) self-concept which includes the minorities as an enriching element and a valuable heritage from earlier days (Translation A.K.).

8   The long-term coexistence of the Czechs with the Austrians within the multinational Habsburg monarchy undoubtedly influenced the customs of both ethnic groups in many ways (Translation A.K.).

The evaluation, that the multinational Habsburg monarchy "undoubtedly" had an influence exposes the statement as the iteration of the myth of the *Habsburg monarchy as a linguistic melting pot*: As a myth, it does not require justification, but is accepted as a true proposition even within the scientific community. In the same volume, Muhr (1995, pp. 226–227) recapitulates several grammatical and pragmatic phenomena of Austrian German. For two phenomena, he briefly mentions that they trace back to Czech or Slovak influences without referring to any study that would prove these assertions.[9] This leads to our hypothesis that even the individual linguistic contact phenomena that commonly (re-)occur in the discourse to distinguish DiÖ from the German standard register in Germany can be treated as myths themselves.

Not only has the myth of the *Habsburg monarchy as a melting pot* entered into and shaped the linguistic discourse on German in Austria. Moreover, that discourse has played a considerable role in the mediation of this myth and connected ideologies that mainly emerged during the 19th century into the 21st century. Recent studies indicate that laymen throughout Austria consider language contact the most important driving factor of language change too (cf. Koppensteiner/Kim, forthcoming).

3.2 The Linguistic Melting Pot throughout the 20th Century in the Light of Dictionaries

Dictionaries of any kind are probably the most widely received linguistic or lay linguistic genre. Hence, they are especially powerful when it comes to the mediation and strengthening of linguistic myths and ideologies. For good reason, Silverstein (1979, p. 193) identifies dictionaries "as the codified authority on what words really mean." However, they do not only codify a language's or variety's lexicon, but also collect "the names of all the distinctive institutions of a culture," which makes them "depositories of the whole culture in microcosm" (Considine 2008, p. 15). In doing so, they considerably contribute to the construction of their object language or variety and the (re-)production of powerful language myths and ideologies.

In this light, the representation of the myth of the *Habsburg monarchy as a melting pot* in dictionaries of the standard variety of German in Austria (or, more

---

9   These phenomena are the possibility of the verb *vergessen* 'to forget' to govern a prepositional argument (cf. Kim/Scharf/Šimko, this volume) and the more frequent use of reflexive pronoun with certain verbs, e.g. *Es lohnt sich nicht* 'It is not worth', *Das geht sich nicht aus* 'It is not enough', etc.

commonly: Austrian German) are of special interest. *The Austrian Dictionary* ("Österreichisches Wörterbuch", ÖWB), which appeared as a school dictionary in 1951 for the first time, pursued and supported a policy of (linguistic) differentiation from Germany (cf. de Cillia/Wodak 2006, p. 38). In its first editions, it does not refer to the Habsburg monarchy at all in its foreword. The revised 38th edition from 1979 (ÖWB 1979), however, identifies it as an important factor in the shaping of an independent Austrian lexicon:

> Zwei Sachbereiche sind es vor allem, in denen die Besonderheiten des österreichischen Wortschatzes deutlich sinnfällig werden: die in der österreich-ungarischen Monarchie entwickelte österreichische Amtssprache und die durch die Wiener Küche unter dem Einfluß mehrerer fremder Nachbarsprachen und -kulturen geschaffene österreichische Küchensprache. (ÖWB 1979, pp. 9–10)[10]

Similarly, the dictionary of Austrian German edited by the prestigious Duden publishing house, Ebner (1st edition: 1969; 4th edition 2009), highlights language contact as a main distinction of the Austrian from the German lexicon (cf. Ebner 1969, p. 254; 2009, p. 444). Both editions identify the 19th century as the crucial period, during which a specific Austrian identity and thus language was developed, e.g.:

> Ein neues habsburgisches Kulturbewusstsein der österreichisch-ungarischen Monarchie sollte staatstragend werden. Erst jetzt entstanden fassbare Ausprägungen eines österreichischen Deutsch, vor allem in der Sprache der Verwaltung, des gesellschaftlichen Lebens und der Koch- und Speisenkultur. (Ebner 2009, p. 440)[11]

Like the ÖWB, Ebner equates language contact with cultural contact and in this context acknowledges the special importance of culinary language:

> Aus dem Zusammenleben der vielen Völker (Deutsche, Tschechen, Slowaken, Kroaten, Italiener, Ungarn, Galizier, Siebenbürger, Slowenen u. a.) entstand eine ganz neue Kultur und Lebensart. Sie fand einen deutlichen Niederschlag in der österreichischen Küche, die aus allen Teilen der Monarchie das Beste zusammengetragen hat. (Ebner 1969: 255)[12]

---

10 There are mainly two areas, in which the particularities of the Austrian lexicon become clearly evident: the Austrian administrative style, which developed in the Austro-Hungarian monarchy, and the culinary language created by the Viennese cuisine which was influenced by several neighbouring foreign languages and cultures (Translation A.K.).

11 A new Habsburgian cultural consciousness was meant to be supportive of the state. Only then did distinct manifestations of Austrian German emerge, especially in the administrative register, the language of high society, and the culinary culture (Translation A.K.).

12 From the coexistence of many nations (Germans, Czechs, Slovaks, Croats, Italians, Hungarians, Galicians, Transylvanians, Slovenes, etc.) a completely new culture and

The identification of linguistic with cultural contact undoubtedly made the myth of the *Habsburg monarchy as a melting pot* an especially powerful one in the construction of a distinct Austrian identity. However, first studies within the SFB "German in Austria" indicate that apart from the names of certain dishes of high popularity in (parts of) Austria, e.g., *Palatschinken*, engl. 'pancakes', *Powidl*, engl. 'plum jam', or *Wuchteln*/*Buchtel*, engl. 'filled sweat rolls', linguistic phenomena that can be traced to the language contact situation within that melting pot are hardly known by speakers today.

3.3 The Linguistic Melting Pot in the Scientific Discourse of the 19th Century

The picture of the *Habsburg monarchy as a linguistic melting pot* with significant influence on various registers of German is, however, not an innovation of the 20th century, and was—at least in the 19th century—geographically not restricted to Austria. In his speech *On an academy of the German language* ("Über eine Akademie der deutschen Sprache"), the rector of the University of Berlin, Emil Du Bois-Reymond,[13] who is considered one of the most influential scientists of the 19th century, argued for the installation of an Imperial Academy of German. Following the example of the *Académie française*, that institution should dedicate itself to the cultivation of German. Amongst others, Du Bois-Reymond (1874, p. 22) identifies a challenge for this aim "in the existence of a second centre of German cultivation in the south-east":

> Spät von der deutschen literarischen Bewegung ergriffen, unter dem Einfluss eines babylonischen Zungengemisches, liess der österreichische Stamm in seiner Sprechweise eine Mengen Eigenheiten sich einwurzeln, welche ebenso schwer zu beseitigen, wie vom classischen Standpunkte zu dulden sind. (Du Bois-Reymond 1874, p. 22)[14]

From Du Bois-Reymond's foreign, explicitly Prussian perspective, DiÖ significantly deviated from German in Prussia or from the "classic" German. Interestingly, Du Bois-Reymond does not trace these differences to dialectal, i.e.,

---

lifestyle arose. This was notably reflected in the Austrian cuisine, which incorporated the best from all parts of the monarchy (Translation A.K.).

13 Emil Du Bois-Reymond (* 7. 11. 1818 Berlin, † 26. 12. 1896 Berlin), cf. Ronge (1959).

14 The Austrian tribe, which was seized by the German literary movement belatedly and which is influenced by a Babylonian mixture of languages, allowed many idiosyncrasies to take root in its variety. These are just as difficult to erase as they cannot be tolerated from the classical point of view (Translation A.K.).

diatopic variation but rather to language contact. However, the author himself was a renowned scholar, but neither a linguist nor a philologist.

Needless to say, contemporary linguists addressed the question of language contact of German in the Habsburg monarchy, too. The first specifically linguistic publication on the matter is Schleicher (1851). August Schleicher[15] published this six page paper titled *On the mutual influence of Bohemian and German* ("Über die wechselseitige Einwirkung von Böhmisch und Deutsch") one year after he had been appointed an extraordinary professorship for comparative linguistics and Sanskrit at the Charles-Ferdinand University in Prague (cf. Bense 2007, p. 50). In the very beginning of the text, he highlights the importance of historical comparative linguistics to consider language contact in their reconstruction of earlier linguistic stages:

> Es ist eine bekannte, bei der vergleichenden Sprachforschung wohl zu berücksichtigende Erscheinung, daß geographisch benachbarte Sprachen, auch wenn sie verschiedenen Familien, ja selbst verschiedenen Stämmen angehören, einen mehr oder minder bedeutenden wechselseitigen Einfluß auf einander üben. (Schleicher 1851, p. 38)[16]

According to Schleicher (1851), mutual influence is the default case in language contact and thus he also describes the contact situation that he daily witnesses in Prague as having influence on "the German language as well as on the Bohemian vernacular" (cf. Schleicher 1851, p. 38–39). The sequence already indicates his focus on German as the target language, which is even clearer in the following passage:

> Und zwar erstreckt sich der Einfluß des Slawischen nicht nur auf das in Böhmen gesprochene Deutsch, sondern auch auf das österreichische, ja in gedruckten Büchern, in Zeitschriften u. dergl. sind Slawismen nicht selten. (Schleicher 1851, p. 39)[17]

This passage can be interpreted as an early scientific recognition of a distinct variant of printed DiÖ, which is shaped by language contact with Slavic languages in general and Czech in particular. Interestingly, Schleicher (1851, p. 39) identifies quite little contact phenomena on the lexical level and claims that the contact

---

15 August Schleicher (* 19. 2. 1821 Meiningen, † 6. 12. 1868 Jena), cf. Bense (2007).

16 It is a well-known phenomenon, which should be taken into account with comparative linguistics, that geographically adjacent languages exert a more or less substantial mutual influence on each other, even if they belong to different families or even stems (Translation A.K.).

17 The influence of Slavic is not only restricted to the spoken German of Bohemia, but extends to Austrian German too. Slavisms are not even unusual in printed books, newspapers and the like (Translation A.K.).

with Czech has influenced German mainly on the syntactic level. Contemporary publications, on the other hand, often list significantly more or even exclusively lexical items, while overtly or covertly reject the possibility of syntactic contact phenomena in the standard variety (cf., e.g., Spačilová 1995; Pohl 1997).

Schleicher (1851, p. 38) refers to the contact situation in Bohemia as "one of the most remarkable examples of mutual influence" and thus emphasises its value as a showcase for general and historical linguistics. His student, the Romance philologist and founder of Creolistics,[18] Hugo Schuchardt[19] assesses research into language contact in the Habsburg monarchy in his seminal publication *Slavic-German and Slavic-Italian* ("Slawo-Deutsches und Slawo-Italienisches", Schuchardt 1884) in a similar way:

> Nirgends findet sich ein günstigerer Boden für Sprachmischung als in unserer Monarchie; zu Beobachtungen nach dieser Seite hin anzuregen ist ein Hauptzweck der vorliegenden Schrift. Es kam mir der Gedanke, das friedliche Wellenspiel näher in's Auge zu fassen welches sich bei dem Zusammenprall deutscher und slawischer Sprachfluth in Cisleithanien erzeugt, […]. (Schuchardt 1884, p. 17)[20]

In contrast to Schleicher (1851), Schuchardt does not limit his observations to any crown land but rather develops an abstract theoretical model of the language contact situation in Cisleithania (cf. Schuchardt 1884, p. 18–21; Fig. 1). He describes four concentric circles, each representing a certain register of German, which is mainly characterised by its speakers. The innermost circle is the German jargon of speakers of Slavic languages and especially Czech, i.e., the German that these speakers acquire in an uncontrolled environment either in the bilingual crown lands or after migrating to a German speaking area.[21] This register is, of course, the one with the most Slavic features on all linguistic levels. As indicated by the arrow in Fig. 1, some of these Slavisms make it through to the second circle, i.e., the German spoken by educated Slavs, bilinguals who
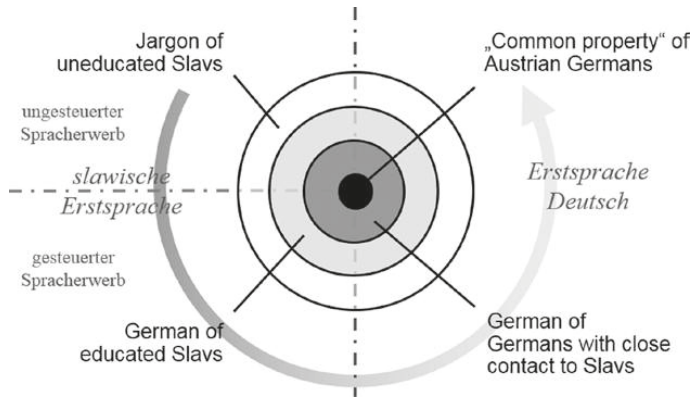
---

18 As such, he strongly disagrees with the idea of a pure language and rather claims that there is no such thing as a completely unmixed, or a completely pure language (cf. Schuchardt 1884, p. 5).

19 Hugo Schuchardt (* 4. 2. 1842 Gotha, † 21. 4. 1927 Graz), cf. Hurch (2007). In 1860, Schuchardt studied at Schleicher in Jena.

20 Nowhere else can a better soil for language mixing be found than in our monarchy. One of the main aims of the present publication is to encourage observations of this kind. I thought about observing the peaceful waves created by the clash of the German and Slavic language floods in Cisleithania, […] (Translation A.K.).

21 This register was and is commonly addressed as *Böhmakeln* or *Kuchldeutsch* and also highly debated and stereotyped throughout the 19th and 20th century.

Jargon of
uneducated Slavs

„Common property" of
Austrian Germans

ungesteuerter
Spracherwerb

*slawische
Erstsprache*

*Erstsprache
Deutsch*

gesteuerter
Spracherwerb

German of
educated Slavs

German of
Germans with close
contact to Slavs

Fig. 1: The concentric circle model of language contact in the Habsburg monarchy according to Schuchardt (1884)

underwent proper foreign language education. They again pass some of their Slavic features on to the monolingual German population in their surroundings. Thus, some of these Slavic features have already become "common property" of the German Austrians.

According to Schuchardt (1884, p. 22), army and administration play a special role in the transmission and propagation of Slavic features in German. He, for instance, describes the administrational register as follows:

> Über der östreichischen[!] Umgangssprache schwebt gleichwie ein wunderbarer Baldachin an welchem Welsche und Slawen in lustiger launiger Weise gewebt haben, die östreichische[!] Kanzleisprache. (Schuchardt 1884, p. 22)[22]

Additionally, Schuchardt (1884, pp. 22–23) emphasises the factor of Jewish migration from the eastern parts of the monarchy to the Austrian crown lands and is convinced, that the specific features in DiÖ, which can be traced back to language contact, are recent developments.

Both Schleicher and Schuchardt represent a distanced perspective on the contact situation and its linguistic consequences in German all over the Habsburg monarchy in two respects. On the one hand, both were born outside of the Habsburg monarchy: at the time when Schuchardt (1884) was published, the

---

22 Similar to a fantastic baldachin woven by the Romance speakers (Welsche) and the Slavs in a jolly and witty manner, the Austrian chancellery language floats above the Austrian colloquial language (Translation A.K.).

author had already held the chair for Romance philology in Graz, and had been living in the Habsburg monarchy for eight years. This leads him to point out, that the differences between the German spoken in the Habsburg monarchy and that in other German countries do not only reflect dialectal differences. They rather make an "un-German impression," i.e., stem from language contact.

On the other hand, both Schleicher and Schuchardt adopt an explicitly scientific perspective, which seeks to gain insight into general mechanisms of language contact and change. Probably, the contemporarily common *völkisch*-nationalistic tone is hence almost absent in Schuchardt (1884), even though the ideology of the *language defining an individual's national identity* and the accompanying ideology of *mixed languages as a threat for an individual's national identity* become evident in some passages, e.g.:

> Andererseits stumpft sich bei Deutschen die lange Zeiträume hindurch fremde Sprachen oder auch nur das Deutsche Fremder um sich hören, das Sprachgefühl nicht in geringem Masse ab; sie nehmen leicht Fremdes an und nicht etwa nur einzelne „verba castrensia" […]. Bei den Söhnen solcher unstäten Väter *wird nun aber selbst die Nationalität streitig*; sie *lernen eine Sprache um die andere*, vergessen auch wohl eine um die andere und es fehlt ihnen, um mit Goethe zu reden, „das Element aus dem die Seele ihren Athem schöpft". (Schuchardt 1884, p. 22; emphasis A.K.)[23]

Contemporary scientists, who were born in the Habsburg monarchy, adopt a different, rather involved point of view. Johann Willibald Nagl,[24] for instance, a German philologist and one of the founding fathers of the Viennese dialectological school,[25] describes *The most important relations between the Austrian and the Czech dialect* ("Die wichtigsten Beziehungen zwischen dem österreichischen und dem čechischen Dialect") in 1887/88, being well aware that he was contributing to a politically sensitive topic. However, he defends himself against accusations from any side by claiming that the linguistic features he deals with date entirely to a time, when the contemporary national conflicts were not yet as

---

23 On the other hand, the feel for language of those Germans who are surrounded by foreign languages or only German spoken by foreigners for a long time wears away. Easily, they adopt foreign elements, and not only single "verba castrensia" […]. For sons of such unstable fathers, *even the nationality becomes controversial*. They *acquire one language after the other*, probably even forget one after the other, and they lack, to cite Goethe, "the element from which the soul draws its breath" (Translation A.K.).

24 Johann Willibald Nagl (* 11. 5. 1856 Natschbach bei Neunkirchen, † 23. 7. 1918 Diepolz bei Neunkirchen), cf. Hornung (1976).

25 Nagl as a dialectologist is exclusively interested in the colloquial, dialectal register.

tangible (cf. Nagl 1887, p. 356). Thus, according to Nagl, they probably did not emerge in the second half of the 19th century.

In contrast to Schuchardt (1884), Nagl (1887) propagates the chauvinistic ideology of *the Germans being spiritually superior to the Slavs*, which according to him is reflected in the according languages. He draws the picture of conservative, independent and determined Germans as opposed to submissive, ruthless, calculating Slavs, who are willing to adapt (linguistically) as soon as it is advantageous for them (cf. Nagl 1887, pp. 358–359).

Theoretically, there is an interesting detail in Nagl's account, namely his early distinction of *Entlehnungen* 'borrowings', or *matter borrowing*, and *Anlehnungen* 'convergence phenomena', or *pattern replication*, as modern contact linguistic theories would frame it (cf. Matras 2010). Most of Nagl's explanations of single phenomena, however, require careful reconsideration.

In the publications of Primus Lessiak,[26] a later representative of the Viennese dialectological school, the bilingual regions of Cisleithania do not provide a peaceful impression but rather that of a "linguistic battleground" (Lessiak 1910, p. 274). Similar to Nagl, this publication is only concerned with dialectal registers, in this case in the German–Slovenian contact area. Lessiak (1910, p. 279) reiterates the ideology of German superiority when he declares Slovene loanwords in German as of special interest for cultural history:

> [Die] Erforschung [des Einflusses der slowenischen auf die deutschen Mundarten] ist von besonderem kulturgeschichtlichem Interesse deshalb, weil bei der Entlehnung aus dem Slow. das Moment der kulturellen Überlegenheit entfällt, das die Übernahme des Fremden auch ohne sachlichen Grund als bloße Modeangelegenheit begünstigt. (Lessiak 1910, p. 278–279)[27]

Even though this sketch only relates to four selected publications, tendencies are evident: In the middle of the 19th century, scientific publications investigate multilingualism and language contact in the Habsburg monarchy from a distanced and neutral perspective. Mainly, they describe it as a showcase for processes of language change, even though contemporary nationalism and the ideology of *language defining an individual's national identity* shine through. In the late 19th

---

26 Primus Lessiak (* 5. 3. 1878 Köttmannsdorf; † 26. 1. 1937 St. Martin bei Klagenfurt), cf. Hornung (1970).

27 The investigation of the influence of the Slovenian dialects on the German dialects is of special interest for cultural history, because the case of borrowings from Slovenian lacks the element of cultural superiority. This element facilitates the adoption of foreign features even without objective reason, simply as a fashion (Translation A.K.).

century, the linguistic discourse in the Habsburg monarchy radicalises and especially members of the Viennese dialectological school propagate the ideology of *German superiority* in their linguistic publications.

3.4 The Linguistic Melting Pot in the Journalistic Discourse of the 19th Century

A specific peculiarity of the discourse on language contact in the Habsburg monarchy throughout the 19th century is the blurred delamination and reversed relation of scientific and journalistic actors in the (re-)production of knowledge. Journalists produce substantial parts of this knowledge, which is subsequently reiterated and thus legitimated by scientists in their publications. In this process, specific journalistic actors become true authorities on various aspects of language, which results in the transformation of their often ideologically shaped personal opinions[28] into items in the dominant discourse archive and ultimately even language myths. This interrelationship shapes the scientific discourse on language contact in Austria until the midst of the 20th century. For example, in his book *Slavic in Viennese* yet another member of the Viennese dialectological school, Walter Steinhauser[29] (1978, pp. 148–154) takes his readers on a time travel back to Vienna around 1900, where they join the journalist Eduard Pötzl[30] for a trip through Vienna. This faked five pages long chapter includes several specifically Viennese contact phenomena that Steinhauser is aware of and explains their etymology in endnotes. Two authentic texts from the chauvinistic anti-Czech satirical weekly newspaper *Figaro* conclude the chapter. To our knowledge, Pötzl hardly used the phenomena and constructions listed in Steinhauser (1978) in his own feuilletons. However, Pötzl was considered a supreme authority on the Viennese dialect in his lifetime and even consulted as an expert in court (cf. Payer 2014, pp. 199–200). Thus, Steinhauser (1978) obviously refers to Pötzl as an authority in order to legitimate his description and observations.

Similarly, Schuchardt (1884) consults journalistic texts as sources. The earliest one, which he is aware of but which he could not retrieve, is a series of feuilletons by Joseph Schön,[31] which appeared in the German newspaper *Bohemia* between

---

28 All texts analysed in this paper are feuilletons or glosses, i.e., journalistic genres in which the authors explicitly express their personal opinions.

29 Walter Steinhauser (* 7. 2. 1885 Wien, † 3. 8. 1980 Wien).

30 Eduard Pötzl (* 17. 3. 1851 Wien, † 21. 8. 1914 Mödling), cf. Lebensaft (1980).

31 Probably Joseph Schön (Schoen, * 14. 3. 1790 Jaroměř-Josefov/Jermer-Josefstadt, † 2. 2. 1838 Písek), cf. Adel/Lebensaft/Reitterer (1995).

February 2nd and March 11th, 1831. In accordance with the title *Etymologic gimmicks* ("Etymologische Spielereien"), the author describes loanwords from any source language in the target languages Czech and German. In this early account, *völkisch*-nationalistic ideologies are not tangible yet. The author rather represents a clear crown land specific patriotism when he refers to Bohemia as his "Czech-German homeland", which is inhabited by "two main tribes of Europe", who speak "two main languages of Europe" (*Bohemia*, 27.02.1831, p. 4).

> Wir […] sind allenfalls der unvorgreiflichen Meinung, daß alle Sprachen die des Nachbars treufleißig benützten, vieles aus der des Fremdlings, mit dem sie in Berührung gekommen, willig aufnahmen, ohne eben dadurch Abkömmlinge desselben zu werden. Wortähnlichkeiten in verschiedenen Zungen zeugen wohl oft für Stammesverwandtschaft, oft auch nur für bloße Bekanntschaft. (*Bohemia*, Feb 27th 1831, p. 4)[32]

In this passage, Schön neutrally describes language contact as an important factor for language change. The myth that *each tribe/nation has its own language* is tangible, because in his text, the author critically deals with the assumption of early historical linguistics that the origin of a tribe can be traced back by comparing and reconstructing its language. The ideologies of *language defining an individual's national identity* and of *mixed languages as a threat for an individual's national identity*, however, are absent in the text, the latter almost denied: according to Schön, language contact does not substantially change a language's and its tribe's or nation's distinct identity.

Two later journalistic contributions, which Schuchardt (1884) refers to and exploits as sources (cf. section 4), are the feuilletons called *Linguistic schmooze* ("Linguistische Plaudereien") by the journalist, writer and theatre director from Prague, Heinrich Teweles,[33] and an essay on the *Abuse of the German language in Austria* ("Misshandlungen der deutschen Sprache in Österreich") published anonymously in the newspaper *Homeland* ("Die Heimat"). Both articles appeared in 1884, i.e., in the same year as Schuchardt published his seminal book, in the

---

32 We are convinced that all languages, which diligently used their neighbour's tongue, and willingly absorbed much of the foreigner's language that they encountered, did not necessarily become descendants of the latter. Similar words in different languages often exemplify tribal relationships, but also frequently merely acquaintanceship (Translation A.K.).

33 Heinrich Teweles (* 13. 11. 1856 Praha/Prague, † 9. 8. 1927 Prein an der Rax), cf. Venus (2014).

preparation of which he exchanged letters with Teweles (cf. Schuchardt 1884, p. 24).[34]

Teweles (1884), a collection of feuilletons with the title *The battle over language* ("Der Kampf um die Sprache"), is shaped by a clearly involved perspective. Originally, the texts were published in the newspaper *Bohemia* from 1883 onwards. Hence, a comparison of Schön's description with Teweles' is especially informative. The author opens his foreword to the 1884 book with a synthesis of the myth that *each tribe/nation has its own language*, which, in his case, is accompanied by a notion of battle:

> Der Styl ist der Mensch und die Sprache ist das Volk. In der Sprache bewahrt das Volk seine ganze geistige und gesellschaftliche Bildung und Entwickelung. Ein Eroberer kann kommen und die Bauwerke niederreißen und die Büchereien verbrennen – wenn er dem Volke nur seine Sprache läßt, so kann es Alles wieder von Neuem aufbauen. In seiner Sprache lebt das Volk wieder auf, und nur wenn es die Sprache verliert, verliert es das Volksthum. (Teweles 1884, p. 1)[35]

The opening statement, an equation of style and the individual human being on the one hand, and language and nation on the other hand, clearly expresses the above-mentioned myth. Teweles, however, goes one step further and even identifies a language—dissociated from the literary products created in that language—as the crucial expression of national traditions. With regard to the situation in Cisleithania, the author explicitly lists the "battlegrounds":

> Und nun zu uns, die wir in Oesterreich, in „Cisleithanien", „in den im Reichsrathe vertretenen Königreichen und Ländern" leben. Wir leben im steten Kampf. Wir haben zu bewahren, was unser und was euer ist. Wir kämpfen um jedes Dorf, um jeden Mann. Die Sprache ist unsere Fahne, sie ist uns geblieben. Bei jeder Schule, bei jeder Amtshandlung vor Gericht, vor der Verwaltungsbehörde, vor dem Gemeindeamt setzt es einen Kampf um die Sprache. (Teweles 1884, p. 5)[36]

---

34 According to information in the online Schuchardt archive, six letters of Teweles can be found in the library of the Karl-Franzens University of Graz (cf. Mattes 2013 and Hurch [Web]).

35 The style is the individual human and language is the nation. In its language, a nation preserves its complete spiritual and societal education and development. A conqueror may come and break down buildings and burn libraries – if he leaves the nation its language, it can rebuild everything from scratch. In its language, the nation is revived and only when it loses its language, it loses its national traditions (Translation A.K.).

36 Now, as for us, who live in Austria, in "Cisleithania", in "The Kingdoms and Lands Represented in the Imperial Council". We live in constant battle. We need to preserve, what is ours and yours. We fight over each village, each man. Language is our banner, it remains ours. In each school, in the course of each official act at court, at

With regard to an important domain of written standard languages, namely to administration, Teweles (1884, pp. 10–11) criticises people, who "build the artistic building of our administrative register" and who claim to master both *Landessprachen*, German and Czech, but actually only speak Czech German or germanised Czech. Throughout the essays, he does not employ a radical anti-Czech tone; however, he frames language mixing as the corruption of the pure language and thus reiterates the ideology of *mixed languages as a threat for an individual's national identity*.

The anonymous essay on the *Abuse of the German language in Austria*, which also appeared in 1884, does not only tackle language contact phenomena, but more generally the "countless small sins" which German speaking Austrians commit "against the spirit or structure of the German language" (cf. w 1884, p. 27). Slavisms, however, are especially highlighted as large groups of mistakes. Not only the fact that the essay refers to contact phenomena as "mistakes", "abuses" or even "sins" indicates its ideological positioning. Furthermore, the author explicitly refers to one of the main publications of the Galician sociologist Ludwig Gumplowicz[37], *The Fight of the Races* ("Der Rassenkampf"), which appeared in its first edition in 1883 (cf. w 1884, p. 28). Gumplowicz, who held a professorship in Graz simultaneously to Schuchardt and who at least received Schleicher's ideas, claims that dialects mainly develop due to language contact:

> Nehmen Fremde eine neue, ihnen durch Umstände und Verhältnisse sich darbietende oder aufgezwungene Sprache an, so werden sie dieselbe nicht so sprechen, wie diejenigen, von denen sie dieselbe annehmen – vielmehr werden sie aus der neu angenommenen Sprache einen Dialekt oder gar, indem sie diese mit Überbleibseln ihrer früheren Sprache vermengen, einen Jargon bilden. (Gumplowicz 1909, p. 129)[38]

The anonymous essay does not only cite these ideas, but also directly reacts to Gumplowicz, since he referred to German in Silesia, his land of origin, as a good example for these processes (cf. w 1884, p. 28). The essay aims to prove this claim with a list of examples. This, of course is not yet an expression of a certain ideological positioning. The part of the text, that deals with Slavisms, however, is

---

    an administrative authority, at the local authority, a battle over language takes place (Translation A.K.).

37 Ludwig Gumplowicz (* 9. 3. 1838 Kraków/Cracow; † 19. 8. 1909 Graz), cf. N.N. (1958).

38 If foreigners adopt a new language that has either presented itself in certain circumstances or under certain conditions or which has been imposed on them, they will not speak it as those [people], who introduced it to them – they rather will create their own dialect from that newly adopted language, or even a jargon, if they mix it with relicts from their own language (Translation A.K.).

illustrated with a copy of a drawing by French painter Hector Giacomelli[39] called *An Intruder* ("Ein Eindringling") which depicts a large grasshopper entering a small bird's, probably a Eurasian penduline tit's, nest. The bird appears to cry in terror. The interaction of the illustration with the text suggests that the essay expresses a generalisation of the ideology of *mixed languages as a threat for an individual's national identity*, namely the ideology of *mixed languages as a threat for the nation as such*. Linguistic awareness and language cultivation thus become important duties of each nationally aware individual. Language contact phenomena and language mixing, on the other hand, carry the notion of threat to the vitality of the nation.

To sum up, this chapter has retraced how the dominant myth that *each tribe/nation has its own language* was enacted in various linguistic ideologies throughout the 19th century, such as the ideology of *mixed languages as a threat for an individual's national identity*, the ideology of *mixed languages as a threat for the nation as such*, the ideology of the *language defining an individual's national identity,* or the ideology of *German superiority*. Before World War II, these ideologies significantly shaped the discourse on language contact in general and the Habsburg monarchy as a linguistic melting pot in particular. After 1945, however, a critique of this ideological set-up was indispensable in the justification of Austria as an independent state. Amongst others, a positively assessed myth of *the Habsburg monarchy as a linguistic melting pot* was construed discursively by seminal publications such as dictionaries, i.e., top-down, in order to support a specific Austrian identity. This narrative simultaneously allows the construction of Austria as a contemporarily monolingual German country and the acknowledgement of historical multilingualism.

4　Single Contact Phenomena as Items in the Discourse Archive and Language Myths

The following chapter focuses on individual (alleged) contact phenomena in both their role as items in the discourse archive of linguistic ideologies connected to the myth of *the Habsburg monarchy as a linguistic melting pot* and as language myths, i.e., the subject of narratives, in their own right. In accordance with the whole paper, this section adapts a primarily metalinguistic approach. However, issues regarding the language system and sociolinguistic aspects are indicated as well.

39　Hector Giacomelli (* 1. 4. 1822 Paris; † 1. 12 1904 Menton).

4.1 Prepositions and the High Risk of Infection

Since the list of (alleged) contact phenomena in DiÖ is quite long and comprises phenomena on all linguistic levels, it is reasonable to restrict the following observations to a limited but coherent and informative sample of such phenomena. For several reasons we chose the domain of prepositional arguments (in the following *prep*-arguments or more generally *prep*-constructions), i.e., a phenomenon of verbal valency on the syntax-lexicon interface. First, this choice is made in accordance with the focus of other contributions to this volume (cf. Gaszewski, this volume; Kim/Scharf/Šimko, this volume), and second, it is motivated by a striking quote from Schuchardt (1884), in which he characterises the domain of prepositions in general as highly prone to language contact induced changes:

> Kaum auf irgend einem Gebiete begeht der Fremde zahlreichere Fehltritte als auf dem der Präpositionen […], und hier lässt sich der Einheimische um so leichter anstecken als ja auch für ihn der richtige Gebrauch derselben nicht immer leicht ist, und sie gern in synonymen und auch formell ähnlichen Verbindungen wechseln. (Schuchardt 1884, p. 115)[40]

These observations correspond to recent findings and theoretical conceptions. Most borrowability hierarchies proposed by modern contact linguistic publications as listed by Matras (2010, p. 76–82; e.g., Thomason/Kaufman 1988; Haugen 1950; Muysken 1981; Moravcsik 1978; Field 2002) describe prepositions, sometimes subsumed in the category of function words, to be of 'medium-high' borrowablity, and thus the category to most likely be borrowed or replicated after content words.

Schuchardt (1884) structures his elaborations on prepositions as contact phenomena in DiÖ by first naming the preposition that occurs in DiÖ and second the preposition that is expected in the contemporary standard register of German. Each such pair is illustrated by at least one example and the example's equivalent in a Slavic language. Tab. 1 explicates the structure of the following quote.

> *Bei* [(1)] für *auf*, so ᴋʟᴀɪᴄ́: *ich verstehe mich nicht beim Einkaufe des Tuches* (*pri kupovanju sukna*). [(2)] Für *um*, so sloweno-d. *es war mir schwer beim Herzen* (*pri srci*). (Schuchardt 1884, p. 116)

---

40 In hardly any other domain, the foreigner makes more mistakes than in the domain of the prepositions […], and in this domain the native tends to get infected easily, since even for him the correct use of the prepositions is not always easy and since they often vary in synonymous and even formally similar constructions (Translation A.K.).

Tab. 1: Structure of Schuchardt's (1884) elaborations on prepositions as contact phenomena

preposition in DiÖ	bei Engl. 'at'	
preposition in the standard register	(1) auf Engl. 'on'	(2) um Engl. 'around, about, at'
register of DiÖ/source	KLAIĆ (= Klaić 1878)	sloweno-d. (= German jargon by speakers of Slovenian)
example from DiÖ	ich verstehe mich nicht beim Einkaufe des Tuches 'I do not know anything about purchasing cloth'	es war mir schwer beim Herzen 'My heart grew heavy'
equivalent in a Slavic language	pri kupovanju sukna (Serbo-Croatian, Engl. 'at purchasing cloth')	pri srci (Slovenian, Engl. 'at heart')

In this manner, Schuchardt (1884, pp. 115–119) lists 118 examples in which the preposition deviates from the contemporary, "classic" German in some register of DiÖ. These *prep*-constructions either correspond to a Slavic construction or, in Schuchardt's opinion, can at least be traced back to the bilingualism of the speakers of DiÖ (cf. the two outermost circles in Schuchardt's concentric circle model, Fig. 1).[41] In many cases and as can be seen in (1) and (2) (cf. Tab. 1), the prepositions in these examples, however, often do not occur in their 'proper', e.g., spatial or temporal function, but rather in more or less fixed constructions. In (1), it is part of a *prep*-argument, in (2) of an idiomatic construction. In Tab. 2, we distinguish between examples with prepositional phrases in adverbials (adjuncts), *prep*-arguments, idiomatic constructions and comparative constructions.

In the following section, the focus lies on the 61 examples for *prep*-arguments, which constitute almost 50 % of all the examples given by Schuchardt (1884), and on (alleged) contact induced variation in DiÖ in the domain of *prep*-arguments in general. Similar to sections 3.3.–3.4., the linguistic discourse from the 19th to the early 20th century is examined in detail in section 4.2., before turning to the extra-linguistic discourse in section 4.3. In both cases, the main goal is to uncover the origin of

---

41 In Schuchardt (1886, p. 347), his additions to Schuchardt (1884), he explicates that he did not intend to suggest total accordance of DiÖ with the Slavic languages by giving the Slavic equivalent of the example in DiÖ in brackets. Instead, he wanted to explicate the deviations from the pattern.

Tab. 2: Examples for *prep*-constructions as contact phe-
nomena in Schuchardt (1884)

	number of examples
adverbials	44
prep-*arguments*	61
idiomatic constructions	11
comparative constructions	2
	118

the idea of certain contact induced *prep*-constructions in DiÖ, their transmission within the scientific as well as the non-scientific community and, thus, the role of these single *prep*-constructions as items in the discourse archive of the discourse on language contact and multilingualism in the Habsburg monarchy and Austria.

4.2 The Discourse on Prep-Arguments as Contact Phenomena in Linguistic Literature

In the analyses of the linguistic discourse on *prep*-arguments as language contact phenomena, we first give an overview of the involved prepositions in DiÖ and their standard German equivalent in order to identify the verbs and constructions in focus. In a second step, the sources consulted by Schuchardt (1884) will be scrutinised in combination with the examples' assignment to a certain variety of DiÖ, which enables us to judge to what extent Schuchardt transmitted already existing ideas of language contact induced changes in DiÖ and to what extend he created them himself. Then, in the last step, the connection to the 21st century linguistic discourse will be made.

Fig. 2 shows the number of examples of *prep*-arguments and—if more examples are given for the exact same construction—the number of verbs, in which a certain preposition in DiÖ (in the rows) corresponds to a different preposition in the (contemporary) "standard register" of German (according to Schuchardt 1884; in the columns).

At first glimpse, Fig. 2 reveals an overrepresentation of *prep*-arguments with the preposition *auf* [+acc][42] 'at, on' in DiÖ, and—according to Schuchardt (1884)—especially in the registers of DiÖ spoken by Slavs (cf. the two outer

---

42 In German, the preposition *auf* 'at, on' belongs to the category of so-called "two-way prepositions", i.e., prepositions which may either govern a nominal phrase in accusative or in dative. In their spatial function, these prepositions generally express directive relations if combined with an accusative, and locative relations in combination with a

preposition in DiÖ (according to Schuchardt 1884)	preposition in the standard variety (according to Schuchardt 1884)																		
	no prep.	an (+acc)	an (+dat)	auf (+acc)	auf (+dat)	aus	für	gegen	in (+acc)	in (+dat)	*mit*	nach	über (+acc)	um	von	vor (+dat)	*zu*	zu/um	
an (+acc)	1			2															3
an (+dat)					1														1
auf (+?)	1																		1
auf (+acc)	2/1	4						1	1		1	1					4/2	2/1	16/12
auf (+dat)			1																1
aus	1									3									4
bei				1															1
durch											3								3
für	2/1										1			3/1					6/3
in (+acc)		1											1						2
in (+dat)																1			1
mit													3		2				5
nach		1															1		2
um						1													1
von	3/2	2				3	1				1					1			11/10
zu	1	1									1								3
	11/8	9	1	3	1	1	2	1	1	3	5	3	3	3/1	2	2	5/3	2	61/53

number of examples/number of verbs

Fig. 2: Language contact induced deviations of DiÖ from the standard variety in the domain of *prep*-arguments according to Schuchardt (1884)

concentric circles in Fig. 1). Schuchardt (1884, p. 115) therefore even calls *auf* the "favourite preposition of the German speaking Slavs".

Visualisations such as Fig. 2 can be consulted to identify system linguistic patterns, i.e., regularities and clusters of phenomena in the domain of *prep*-arguments (allegedly) influenced by Slavic languages in DiÖ. In four of the 16 examples given by Schuchardt (1884), *auf* [+acc] corresponds to standard German *prep*-arguments with the preposition *an* [+acc] 'at, on', all of which contain different verbs (3)–(6).

(3)   der Mond kümmert sich nicht, wenn der Hund <u>auf ihn</u> *bellt* (de-RS)[43]
      'the moon does not care, if the dog barks at it'

---

dative. Throughout this article, these two-way prepositions are always mentioned with the case of the nominal phrase governed by them. In cases, in which it is not possible to determine the case based on the analysed examples, a question mark is added.

43   In the given examples, the verb is typeset in italics and the *prep*-argument underlined. In brackets, the variety/register of DiÖ, that Schuchardt (1884) himself assigns it to, is given according to the scheme of language tags developed in Kim/Breuer (2017). Such a tag basically comprises an ISO 639-1 language code in its first position, in which the given case of course is always *de* 'German'. In the second position, an ISO 3166-1

(4)   <u>auf jemanden</u> *schreiben* (de-CS)
      'to write somebody [a letter]'

(5)   sich <u>auf etwas</u> *erinnern* (de-AT)
      'to remember something'

(6)   i *glaub* <u>auf oan [sic!] Gott</u> (de-SI-x-sor)[44]
      'I believe in God.'

In examples (3) and (4), the *prep*-arguments express the semantic role *addressee* with verbs of communication. In examples (5) and (6), the verbs can be classified as cognitive verbs with the *prep*-arguments expressing the *patiens*-role. Note, that in contemporary standard German, example (3) would rather be formulated with a morphologically complex verb *anbellen* 'to bark at somebody', cf. (7):

(7)   […], wenn <u>ihn</u> der Hund *anbellt*
      'if the dog barks at it'

A comprehensive system linguistic overview remains to be given. In this context, we focus more on the establishment of the single phenomena as contact phenomena in the (extra-)scientific discourse. Thus, the question arises, whether Schuchardt (1884) collected these phenomena himself or whether and which previous literature he received and transmitted.

Fig. 3 presents Schuchardt's (1884) examples including a *prep*-argument according to the varieties of DiÖ to which he assigned them and the sources that he either overtly or covertly quotes them from. So far, we have been able to identify the sources of 48 of his 61 examples for *prep*-arguments as contact phenomena,

---

country code is used to identify the area in which it is spoken/used. Of course, the inadequacy of these codes with regard to historical entities needs to be accepted. If the example is assigned to a certain region, town or village, the language tag is extended by *-x-* and a three letter code to refer to this place. In this article, the abbreviations *sor* (Sorica/Zarz), *vie* (Wien/Vienna), *cer* (Černivci/Czernowitz), *lit* (Litoměřice/Leitmeritz) are used. Another abbreviation *jew* 'jewish' identifies the example as used in a German variety with Jewish/Yiddish impact.

44 Note, that this example is quoted in the dialect of Sorica/Zarz.

varieties	Berlić 1854	Bernd 1820	Burian 1843	Cigale 1860	Czoernig 1875	Ebert 1833	Halatschka 1883	Heinrich 1875	Jungmann 1835–1839	Klaić 1878	Krauss 1883	Politik	Steinsberg 1797	Teweles 1884	w 1884	identified	not identified	overall
de-AT			1												1	2	3	5
de-AT-x-vie								1								1	0	1
de-CS						1	2	1	2		1		4	3		14	3	17
de-HR								2								2	0	2
de-PL		17					1								2	20	3	23
de-RS	5															5	0	5
de-SI				1				1								2	3	5
de-SI-x-sor					1											1	0	1
de-XX																0	1	1
de-YU										1						1	0	1
	3	17	1	1	1	1	4	2	2	2	1	1	3	3	3	48	13	61

Fig. 3: Examples for language contact induced *prep*-arguments in Schuchardt (1884) according to variety and sources

i.e., a very high percentage of 79 %. In all of these cases, Schuchardt (1884) literally copied the examples from the sources, in only 23 of them, however, overtly (i.e., 47 % of the identified ones). First of all, to identify the sources, the overt citations were verified. Secondly, all currently available sources listed in Schuchardt (1884, pp. 18–25) were checked to determine if they contained the examples listed by Schuchardt (1884) or not. To a large degree, this approach was enabled by the fact, that the Austrian National Library (ÖNB) has already digitised a large share of its historical, copyright-free stock of books in the *Austrian Books Online* initiative (cf. ÖNB 2013), which can be accessed and searched online via the online catalogue of the library.[45] Some of the 13 examples, which we were not yet able to identify, are possibly quoted from not (digitally) accessible sources or parts of Schurchardt's correspondence that has not been edited by the *Schuchardt archive* (Hurch [Web]).

The listed sources belong to various types of publications. These are:

a) *Linguistic descriptions* of Slavic languages (Berlić 1855) or varieties of German (Bernd 1820),

b) *Textbooks* of Slavic languages (Burian 1843; Klaić 1878) or of German (Heinrich 1875),

c) *Dictionaries* of Slavic languages (Cigale 1860; Jungmann 1835–1839),

d) *Ethnographic and linguistic publications* with scientific aspiration (Czoernig 1875; Halatschka 1883; Krauss 1883),

---

45 Cf. https://onb.ac.at/ (17. 10. 2018).

e) *Literary texts* (Ebert 1833; Steinsberg 1797), and

f) *Journalistic texts* (Teweles 1884; w 1884; and the *Politik* 'Politics' a rather liberal newspaper published in Prague)

The following interpretation of Fig. 3 focuses on the two varieties of DiÖ that Schuchardt (1884) assigns most of his examples for *prep*-arguments influenced by language contact to, namely German spoken by Czechs, i.e., in Bohemia, Moravia and Austrian Silesia (de-CS, 17 examples), and by Poles, i.e., in what is present day Poland (de-PL, 23 examples). The sources of three examples assigned to each of these varieties could not be identified. However, the difference in number of sources that the rest of the examples stem from is evident:

For the German jargon in what is nowadays Poland, Schuchardt (1884) mainly refers to one source, i.e., Bernd (1820), a linguistic description of German as spoken in the Grand Duchy of Posen (category a), a territory which was never part of the Habsburg empire. Schuchardt (1884, p. 25) is well aware of the fact that he thus partly describes varieties of German spoken in the German empire. Of the 17 examples taken from Bernd (1820), he overtly marks only four. Two examples stem from the anonymously published essay *On the Abuse of the German language in Austria* (w 1884; category f) already dealt with in section 3.4. of this article. Another very instructive example is apparently taken from Halatschka's (1883, p. 32) critical description of Newspaper German (category d):

(8) dass ich <u>von jedem Grusse</u> *vergass* (de-PL)[46]
    'that I forgot each greeting'

The interpretation of this *prep*-construction as resulting from Slavic influence, however, stems from Schuchardt himself. He mainly argues with the author's descent: Johannes Friedrich Meissner[47] was born and grew up in Pomerania (cf. Meißner 1974). The biographical lexicon (Meißner 1974) does not give sufficient information on Meissner to judge, whether the phenomenon shown in (8) would have to be classified as "German of educated Slavs" or "German by Germans with close contact to Slavs" (cf. section 3.2.).

The 13 examples assigned to German spoken in the lands of the Bohemian crown originate from a greater variety of sources and have mainly been quoted

---

46 In the standard register of German, the verb *vergessen* 'to forget' may only govern an *acc*-argument, but never a *prep*-argument. For a broader discussion cf. Kim/Scharf/Šimko (this volume). In Figure 2, this example is thus listed in the first column (no prep.).

47 Johannes Friedrich Meissner (* 25. 2. 1847 Rathsdamnitz/Dębnica Kaszubska; † 4. 3. 1918 Wien/Vienna)., cf. Meißner 1974.

overtly (eight of 13 examples, i.e., 62 %). Furthermore, most of the examples are either literary (category e) or journalistic texts (category f). Two examples are (overtly) taken from Halatschka (1883), one from a textbook of German designed for education in bilingual crown lands (Heinrich 1875) and two from Jungmann's (1835–39) seminal Czech–German dictionary.

The latter is a good example for a general approach of Schuchardt, who seems to have read and analysed a large amount of textbooks (e.g., Burian 1843; Klaić 1878), grammars (e.g., Berlić 1854) and dictionaries (e.g., Cigale 1860; Jungmann 1835-1837) of Slavic languages written by "educated Slavs", i.e., bilinguals with Slavic L1 and an excellent command of their L2 German. From the interpretation of Figure 3 we may judge the role of Schuchardt in the (re-)production of the scientific knowledge about Slavic contact phenomena on DiÖ as crucial in two ways. First, he recited and transmitted many contact explanations for linguistic phenomena in DiÖ, which were contemporarily discussed in the (extra-) scientific discourse, and, second, he added his own observations.

In comparing his examples for German spoken by Czechs and by Poles, it strikes the eye that Schuchardt (1884) lists two different *prep*-argument constructions with the verb *vergessen* 'to forget', namely (8) and (9):

(9)   dass ich <u>auf den Tölpel und Esel</u> *vergass* (de-CS, source: Steinsberg 1797,
       p. 75)
       'that I forgot about the fool and the donkey'

Fig. 4 reveals that across selected literature from the 19th century, the *prep*-arguments with the prepositions *auf* 'on, at' and *von* 'from' are the most common ones with regard to which DiÖ differs from the contemporary German standard. The examples (8) and (9) fit into an areal pattern. Whereas examples with the preposition *von* 'from' mostly stem from Polish-German or even Estonian-German contact areas and thus from the north-eastern border of the Habsburg monarchy and Prussia, examples with the preposition *auf* 'on, at' originate from the Czech–German bilingual regions and thus from the central part of the monarchy. Interestingly, in most cases both prepositions replace a *prep*-construction with standard German *an* 'on, at' (cf. Fig. 2). This areal perspective on certain constructions needs to be examined in future studies.

Due to the political developments in the 20th century—the end of the Habsburg monarchy and the establishment of the Austrian republic in its contemporary borders—it is not surprising that phenomena from the north-eastern border of the former monarchy were not mediated scientifically as specifically Austrian contact phenomena throughout the 20th century into the 21st century. Fig. 4

preposition in DiÖ according to the literature

	an (+acc)	an (+dat)	auf (+?)	auf (+acc)	auf (+dat)	aus	bei	durch	für	in (+?)	in (+acc)	in (+dat)	mit	nach	über (+acc)	um	von	zu	
Schleicher 1851				1											1				2
de-CS															1				2
w 1884			1	1		2							1			1			6
de-AT				1		2										1			4
de-PL			1																1
de-PL-x-cer													1						1
Teweles 1884	1			4													1		6
de-CS	1			4													1		6
Schuchardt 1884	3	1	1	16	1	4	1	3	6		1	1	5	2		1	9	3	58
de-AT				2		2										1			5
de-AT-x-vie	1																		1
de-CS	2			8				3			1						1	1	16
de-HR		1					1												2
de-PL			1	2	1	1			3			1	3	2			7	2	23
de-RS				2									1						3
de-SI						1			3								1		5
de-SI-x-sor				1															1
de-XX													1						1
de-YU				1															1
Schuchardt 1886				4						1	1					1	5		12
de-AT				2							1								3
de-CS				2															2
de-CS-x-lit																1			1
de-EE																	4		4
de-PL-x-jew										1									1
de-XX																	1		1
Zeman 2003				4		2													6
de-AT				4		2													6
Newerkla 2013a				4		1													5
de-AT						1													1
de-AT-x-vie				4															4
	4	1	2	34	1	9	1	3	6	1	2	1	6	2	1	3	15	3	98

Fig. 4: Language contact induced *prep*-arguments in DiÖ in scientific literature from the 19th to the early 21st century

illustrates this aspect too. Since we so far lack a comprehensive overview of contact phenomena specific for DiÖ, the figure only includes the two publications that list more than one *prep*-construction.[48] Additionally, we have hardly been able to find investigations into such phenomena except for some methodically

---

48 For a closer discussion of the literature on a specific *prep*-construction, namely with the verb *vergessen* 'to forget' see Kim/Scharf/Šimko, in this volume. Blahak (2015) also lists more, however he only analyses Franz Kafka's idiolect.

not sophisticated self-report surveys on the usage of, e.g., *vergessen auf* [+acc] 'to forget about' (cf. for a more detailed description Kim/Scharf/Šimko, this volume). However, publications as Muhr (1995) and Zeman (2003) suggest a mainly oral scientific discourse and an according transmission of a certain list of alleged contact phenomena throughout the 20th century. With regard to *prep*-constructions, the latter publication, for example, quotes several personal consultations with German dialectologists in Vienna some of whom opt for the contact explanations, while others are rather against them. This pattern gives evidence for the integration of contact phenomena into the linguistic ideology of a *specifically Austrian German*.

Besides the metalinguistic patterns of transmission, the transmitted linguistic constructions deserve attention. The lists given in Zeman (2013, pp. 275–279) and Newerkla (2013, pp. 252, 255) are basically congruent.[49] Next to the cognitive verbs already mentioned, i.e., *sich erinnern* 'to remember' (5) and *vergessen* 'to forget' (8) they comprise a third one:

(9)   a.   DiÖ.
          <u>auf etwas</u> *denken*
          'to think about something, to remember something'

      b.   Standard German
          <u>an etwas</u> *denken*
          'to think about something, to remember something'

Equally to *sich erinnern* 'to remember' but differently from *vergessen* 'to forget', *denken* in the specific meaning 'to think about something, to remember something' also requires a *prep*-argument in Standard German. In DiÖ, the preposition may differ as in (5) or (9a). However, these constructions are considered substandard.[50] With cognitive verbs, the *prep*-argument expresses the semantic role *patiens*, i.e., the cognitive content that is remembered, forgotten or thought about.

---

49 Both authors have published more than the articles referred to in Figure 4 on the topic (cf. Newerkla 2007, pp. 39–40; Newerkla 2009, p. 10; Newerkla 2013b, pp. 9, 11; Newerkla, this volume; Zeman 2009; Zeman 2011, p. 60).

50 The *Digital dictionary of German* (Digitales Wörterbuch der deutschen Sprache) marks the according constructions as "regional, (especially) Austrian" (cf. DWDS [Web]).

Besides these cognitive verbs, both publications contain a construction with the noun *Vorbereitungen* 'preparations' (Pl.), the valency of which goes back to the verbal valency of the according verb *vorbereiten* 'to prepare' (10). In this case, the *prep*-argument realises the semantic role *aim*, which is also required in Standard German, where the preposition of choice is *für* 'for'.

(10) a. DiÖ.
      Vorbereitungen <u>auf etwas</u>/jemanden <u>auf etwas</u> *vorbereiten*
      'preparations for something'/'to prepare somebody for something'

  b. Standard German
      Vorbereitungen <u>für etwas</u>/jemanden <u>für etwas</u> *vorbereiten*
      'preparations for something'/'to prepare somebody for something'

The last *prep*-construction listed by both Zeman (2003) and Newerkla (2013) does not contain the preposition *auf* 'on, at' but the preposition *aus* 'from' and showcases, that in many instances, the choice of a certain preposition in DiÖ is motivated by narrow semantic restrictions of the noun in the *prep*-argument. The examples in (11) stem from various publications illustrated by Fig. 4. In all of them, the noun of the *prep*-argument refers to a school subject. In the standard register as used in Germany, only the preposition *in* 'in' would be possible in all of the examples. Czech, however, also uses the preposition equivalent to German *aus* 'from', i.e., *z* 'from' (cf. Newerkla 2013, p. 252).

(11) a. de-AT (w 1884, p. 28)
      er hat *seine Prüfung* <u>aus der Mathematik</u> gut *abgelegt*
      'he passed the exam in mathematics with great success'

  b. de-AT (w 1884, p. 28)
      er hat *seine Maturitätsprüfung* mit Auszeichnung <u>aus Latein</u> *abgelegt*
      'he passed his high school exam in Latin with distinction'

  c. de-AT (Schuchardt 1884, p. 116)
      <u>aus etwas (einem Fache)</u> *prüfen*
      'to examine [somebody] in something'

d. de-AT (Schuchardt 1884, p. 116)
<u>aus Italienisch</u> *Unterricht erteilen*
'to give lessons in Italian'

e. de-AT (Zeman 2003, p. 278)
eine Prüfung <u>aus Biologie</u> *bestehen/ablegen*
'to pass/take an exam in biology'

f. de-AT (Zeman 2003, p. 278)
ein „Sehr gut" <u>aus Mathematik</u> *bekommen/kriegen*
'to get/receive a distinction in mathematics'

g. de-AT (Newerkla 2013, p. 252)
eine Prüfung <u>aus Russisch</u> *ablegen*
'to take an exam in Russian'

Interestingly, both the publications from the 19th and those from the 21st century assign the examples given in (10) to DiÖ and its standard register in general and do not limit it to a certain region. Newerkla (2013, p. 525), however, classifies it as outdated. Furthermore, in all the given examples, the verbs are used in a very narrow and fixed meaning. They thus show a high degree of idiomacity and require an argument to express the semantic role *source*, i.e., the school subject.

We are convinced, that such detailed and comprehensive information on the description of contact phenomena in DiÖ are valuable prerequisites for detailed synchronic, diachronic, contrastive, and areal investigations into individual phenomena. As shown above, we so far lack sufficient examinations, even though knowledge on these (alleged) contact phenomena is frequently (re-)produced by scientific publications when it comes to supporting the ideology of a *distinct Austrian German*.

4.3 The Extra-Linguistic Discourse and Prepositional Arguments

Besides being important pieces in the discourse archive of the linguistic ideology of a *specific Austrian German* up until nowadays, individual (alleged) contact phenomena in general and *prep*-constructions in particular constituted language myths themselves, i.e., were subject to "communally shared stories" at least in the Habsburg monarchy. They served as stereotypes and were used to trigger certain evaluations. However, in some cases these evaluations changed after the end of

the Habsburg monarchy and even led to an exploitation of some contact phenomena for terminological purposes.

Illustrative examples for stories which evolved around *prep*-constructions can be given for *vergessen* 'to forget' with a *prep*-argument. So far, only the *prep*-argument with the preposition *auf* 'on, at' has been mentioned, however, until the midst of the 20th century another one with the preposition *an* 'on, at' was similarly wide spread if not wider.[51] Teweles (1884) prominently refers to both constructions in his critique of the bureaucratic language of Prague in particular and Austria in general. He calls the readers' attention to these constructions and reminds them of their alienness, their non-German character. This can be interpreted as the intentional *creation of saliency*: Phenomena, which—the author fears—are not generally recognised as contact phenomena, but which he judges as such, are made identifiable and are even turned into markers for the *mixed*, the *corrupt language*.

> Das Schlimmste daran ist, daß diese verdorbene Sprache uns allgemach ganz vertraut klingt, daß wir oft darauf oder daran vergessen (das ist auch eine jener čechisierten Wortfügungen), daß diese oder jene Wendung, dieses oder jenes Wort gar nicht deutschen Ursprungs ist und nur ein örtliches Verständnis findet. (Teweles 1884, p. 11)[52]

A similar language purist approach can be observed in texts by Karl Kraus[53] in his newspaper *The torch* ("Die Fackel") published before World War I. In the quote below, he uses several *prep*-arguments[54] as markers for the *corrupt language*, which he observes and criticises in contemporary newspapers throughout Vienna:

> Nun wird der Jargon in der Wiener Redaktion bald obligat sein. […] Man kann jetzt die Wiener Journalistik in zwei Gruppen einteilen: die eine, die etwas auf wem weiß, es aber

---

51 Until its 35th edition from 1979 the ÖWB listed the *prep*-argument with *an* 'on, at' first and the one with *auf* 'on, at' second. Nowadays, the first is not present anymore (cf. ÖWB 1951; ÖWB 1979; ÖWB 2016).

52 The worst part about it is the fact that this corrupt language gradually sounds completely familiar to us. Therefore we often forget [about it] (that is also one of these Czech constructions), that one or another of the expressions, one or another of the words is not at all of German origin and only locally understood (translation A.K.).

53 Karl Kraus (*28.4.1874 Jičin/Jičín, Böhmen; † 12.6.1936 Wien/Vienna), cf. Schick (1968).

54 In this case these constructions are: *etwas auf jemanden wissen* instead of *etwas über jemanden wissen* 'to know something about somebody', *etwas auf jemanden sagen* instead of *etwas über jemanden sagen* 'to say something about somebody', and *an etwas vergessen* instead of *etwas vergessen* 'to forget something'.

nicht sagen will, und die andere, die etwas auf wen sagen könnte, aber daran vergessen
hat. (*Die Fackel* 216, 1. 9. 1907)[55]

The stories, i.e., etymologies Kraus proposes to explain these phenomena, how-
ever, differ from the stories told by Teweles (1884): Whereas the latter traces
both the *prep*-argument with *auf* 'at, on' and *an* 'at, on' to language contact with
Czech, Kraus connects the construction *vergessen an* [+acc] to Yiddish or the
Jewish jargon. Both etymologies reoccur in scientific publications. Muhr (1995,
p. 226), Zeman (2003, pp. 275–279) and Newerkla (2013, p. 255) favour Czech
origin. Kretschmer (1918: 7), on the other hand, describes *vergessen auf* [+acc] as
a native German construction and, similar to Kraus, traces *vergessen an* [+acc] to
Jewish (for a closer discussion cf. Kim/Scharf/Šimko, this volume).

Kraus' position towards and evaluation of the phenomenon in question,
vergessen an [+acc], remarkably changes after World War I.[56] Whereas in ear-
lier commentaries like the one quoted above, it is treated as a violation against
German grammar and as a salient marker of the *corrupt language*, he later
describes it as a construction with a meaning of its own right. In his opinion, the
preposition *an* 'on, at' signifies that the process of forgetting sticks *at* the object,
since it is not actually deleted from the memory, but still *accessible* to the actor.
This description strikingly corresponds to the semantics of the construction
described by Kim/Scharf/Šimko (this volume). The following passage can be
read as the recognition of a linguistic construction that was saliently percieved
to be specifically Austrian. The dimension of language contact is not referred
to and the focus seems to be on the construction of a specific linguistic iden-
tity. Whether these observations fit into a larger picture of the developement of
Karl Kraus's linguistic ideologies in the nationalist-fascist political climate of the
inter-war-period requires future research.

Doch dürfen sie darum wieder nicht glauben, daß es unter allen Umständen falsch
wäre. […] Es ist von „sich daran erinnern" oder „daran denken" bezogen, dessen
Neigung nicht zu Ende gedacht ist, so daß aus der positiven Sphäre das „an", das ja
mit der Erinnerung vor allem entschwunden sein sollte, übrigbleibt. So ließe sich der
Fall denken, daß ein „Vergessen", in dem dieser Vorgang noch sehr stark betont sein

---

55 Soon, the jargon will be obligatory in the Viennese editorial department. […]
    Nowadays, Viennese journalism can be categorised into two groups: the one that knows
    something on [sic!] somebody, but does not want to tell, and the other that could tell
    something on [sic!] somebody, but forgot about it (Translation A.K.).

56 This assessment is based on an analysis of all hits for the infinitive *vergessen* 'to forget'
    in the online Fackel-corpus of the Austrian Academy of Sciences (cf. Austrian Academy
    Corpus 2018 [Web]).

möchte, etwa mit jener Absichtlichkeit, die sich nicht erinnern will, noch „an" dem Objekt haften bliebe. […] Man könne von einem unzuverlässigen Zeugen, der sich an etwas nicht erinnern kann, woran er sich nicht erinnern will, wirklich sagen, er habe „daran vergessen" und man hätte dem psychischen Sachverhalt keinen Abbruch getan. (*Die Fackel* 572–576, 6/1921)[57]

Karl Kraus was regarded as an important authority for language use during his lifetime. Thus, it is not surprising, that in consistency with and even with reference to Karl Kraus's semantic description of *vergessen an* [+acc], a few years later Adolf (Albert) Storfer[58] proposes a clear definition of both the *prep*-argument and the *acc*-argument construction for terminological purposes in psychoanalysis. His attempt particularly intends to justify and explain what had been identified as a "special character" in Sigmund Freud's œuvre by earlier observers (Storfer 1932, p. 364–365). His definitions read as follows:

an etwas vergessen (= etwas, was nicht verdrängt ist, also aus dem Vorbewußten reproduziert werden könnte, nicht reproduzieren, weil die Reproduktion selbst irgendwie vereitelt wird)
etwas vergessen (= etwas ins Unterbewußte verdrängt haben, es daher nicht reproduzieren können) (Storfer 1932, p. 369)[59]

Similar to section 3.3 with regard to the myth of the *Habsburg monarchy as a linguistic melting pot*, this chapter has exemplified the close interaction of the journalistic and the linguistic and other scientific discourses in the production and reproduction of language myths about individual contact phenomena. So far, these examples do not allow for a judgment with regard to whether these stories

---

57 However, we must not judge it wrong under all circumstances. […] It derives from "sich daran erinnern" (to remember something) or "daran denken" (to think about something), the tendency of which is not consequently thought through. Thus, the "an" (at, on) remains from the positive sphere even though it should have vanished together with the memories. Hence one could think of a case in which the "forgetting" stuck on the object, since that process is emphasised strongly, e.g., with the intention to not remember the object. […] One could truly say about an untrustworthy witness who cannot remember something he does not want to remember, that he forgot about it and would not infer with the psychological circumstances. (Translation A.K.)

58 Adolf Joseph (Albert) Storfer (* 11. 1. 1888 Botoșani/Botoschan, Romania; † 2. 12. 1944 Melbourne, Australia), cf. Venus (2009).

59 to forget about something (= to not reproduce something not suppressed that could therefore be reproduced from the pre-consciousness, because the reproduction itself is somehow disabled)
to forget something (= to have suppressed something into one's sub-consciousness and therefore not be able to reproduce it) (Translation A.K.)

emerged in the extra- or intrascientific context. However, they indicate that these phenomena were subject of discussion in several domains of society in the late 19th and early 20th century.

5 Summary and Conclusion

This article has served two main, closely interconnected purposes. First, it has given an insight into the methods of the project part "German and the Slavic languages in Austria: Aspects of language contact" of the SFB "German in Austria: Variation – Contact – Perception". This project part ultimately aims at identifying language myths regarding the contact of German in Austria and Slavic languages.

Second, in applying these methods, it could be shown, how individual language contact phenomena both serve as items in the discourse archive of the myth of the *Habsburg monarchy as a linguistic melting pot* while simultaneously revealing the character of language myths themselves. The following results deserve to be highlighted:

- The article has argued that throughout the Second Austrian Republic, the myth of the *Habsburg monarchy as a linguistic melting pot* has been discursively construed top-down in order to support a specific Austrian identity. It plays a crucial role in the definition of Austria as a monolingual German nation-state rather than a part of a German 'cultural nation'.
- However, it incorporates both discursive patterns and linguistic phenomena as elements in the discourse archive that had already been decisive within the language myth dominant throughout the 19th century, namely the myth that *each tribe/nation has its own language*.
- In a subsequent step, *prep*-arguments were chosen as examples for linguistic phenomena to show the interconnectedness of the journalistic and the scientific discourse in the 19th and early 20th century. Both contributed equally to the production and the reproduction of myths, i.e., stories and etymologies of alleged linguistic contact phenomena. Similar to the more general language myths that relate to mono- and multilingualism, we can observe a change in focus and assessment in the early 20th century, which was completed after the end of World War II.

This article has also argued for the value of detailed metalinguistic evaluation of linguistic and extra-linguistic descriptions of Slavic–German language contact. Even though these mostly consist of lists of alleged contact phenomena, their evaluation reveals linguistic and areal patterns necessary for the design

of detailed studies. Such studies, however, are indispensable in the plausibility assessment of contact explanations and thus in the identification of language myths regarding individual linguistic phenomena of German in Austria.

References

Austrian Academy Corpus: *AAC-FACKEL. Digital Edition No 1*. Retrieved at 8. 8. 2018, from http://www.aac.ac.at/fackel.

Adel K./Lebensaft, Elisabeth/Reitterer, Hubert: "Schön (Schoen), Josef (1790-1838), Schriftsteller, Pädagoge und Priester." In: *Österreichisches Biographisches Lexikon 1815‒1950*, Bd. 11, Lfg. 51, 1995, pp. 36–37.

Bense, Gertrud: "Schleicher, August". In: *Neue Deutsche Biographie* 23, 2007, p. 50.

Berlić, Andreas-Torquat: *Grammatik der illyrischen Sprache wie solche im Munde und Schrift der Serben und Kroaten gebräuchlich ist*. Wien: Mechitaristen-Buchdruckerei 1854.

Bernd, Christian Samuel Theodor: *Die deutsche Sprache in dem Großherzogthume Posen und einem Theile des angrenzenden Königreiches Polen*. Bonn: Eduard Weber 1820.

Blahak, Boris: *Franz Kafkas Literatursprache. Deutsch im Kontext des Prager Multilingualismus*. Köln: Böhlau 2015.

Budin, Gerhard et al.: "Der Spezialforschungsbereich 'Deutsch in Österreich (DiÖ). Variation – Kontakt – Perzeption'. The Special Research Programme 'German in Austria (DiÖ). Variation – Contact – Perception'". In: *Zeitschrift für germanistische Linguistik* 46(2), 2018, pp. 300–308.

Bundesverfassungsgesetz. In: Bundesgesetzblatt 1/1920. Retrieved at 18. 10. 2018 from http://alex.onb.ac.at/cgi-content/alex?aid=bgb&datum=19200004&seite=00000001.

Bundesverfassungsgesetz. In: Bundesgesetzblatt 68/2000. Retrieved at 18. 10. 2018 from https://www.ris.bka.gv.at/Dokumente/ BgblPdf/2000_68_1/2000_68_1.pdf.

Burian, Tomáš: *Ausführliches, theoretisch-praktisches Lehrbuch der böhmischen Sprache für Deutsche. 2. Auflage*. Prag, Königsgrätz: Pospissil 1843.

Cigale, Matej: *Deutsch-slovenisches Wörterbuch. Nachdruck der Klagenfurter Ausgabe*. Klagenfurt, Wien u. a.: Wieser 2004 [originally: Laibach 1860].

Considine, John: *Dictionaries in Early Modern Europe. Lexicography and the Making of Heritage*. Cambridge: Cambridge University Press 2008.

Czoernig, Karl Freiherr von: "Aus dem oberen Isonzo Gebiete. I Idria. II Die vergessene deutsche Sprachinsel Deutschruth. III Auf den Matajur. Vortrag

gehalten in der Section Küstenland des Deutschen und Oesterreichischen Alpenvereins am 29 October 1874". In: *Zeitschrift des Deutschen und Oesterreichischen Alpenvereins* VI, 1875, pp. 243–256.

de Cillia, Rudolf/Wodak, Ruth: *Ist Österreich ein „deutsches" Land? Sprachenpolitik und Identität in der Zweiten Republik* (Österreich – Zweite Republik. Befund, Kritik, Perspektive 16). Innsbruck, Wien, Bozen: Studienverlag 2006

Du Bois-Reymond, Emil: *Über eine Akademie der deutschen Sprache. Über Geschichte der Wissenschaft. Zwei Festreden in öffentlichen Sitzungen der Königl. Akademie der Wissenschaften zu Berlin*. Berlin: Ferd. Dümmlers Verlagsbuchhandlung 1874.

DWDS 2018 = Berlin-Brandenburgische Akademnie der Wissenschaften (ed.): *DWDS - Digitales Wörterbuch der Deutschen Sprache*. Retrieved at 17. 8. 2018, from https://www.dwds.de/.

Ebert, Karl Egon [alias: Wokraulitschek, Franta]: *Böhmische Kolatschen. Eine Sammlung böhmischer Charakterzüge und lustiger Anekdoten*. Leipzig: Wigand 1833.

Ebner, Jakob: *Duden – Wie sagt man in Österreich? Wörterbuch der österreichischen Besonderheiten*. Mannheim, Wien: Dudenverlag 1969.

Ebner, Jakob: *Duden – Wie sagt man in Österreich? Wörterbuch des österreichischen Deutsch*. 4., völlig überarb. Aufl. Berlin, Wien: Dudenverlag 2009.

Ernst, Peter: "Die Tschechen in Wien und ihr Einfluss auf das Wienerische. Eine kritische Bestandsaufnahme". In: Nekula, Marek/Bauer, Verena/Greule, Albrecht (eds.): *Deutsch in multilingualen Stadtzentren Mittel- und Osteuropas um die Jahrhundertwende vom 19. zum 20. Jahrhundert*. Wien: Praesens 2008, pp. 99–107.

Federal Constitutional Law. Official translation, retrieved at 3. 4. 2018, from https://www.ris.bka.gv.at/Dokumente/Erv/ERV_1930_1/ERV_1930_1.pdf.

Ferguson, Charles A[lbert]: "Diglossia". In: *Word* 15, 1959, pp. 325–340.

Field, Federic W.: *Linguistic Borrowing in Bilingual Contexts*. Amsterdam: John Benjamins 2002.

Glauninger, Manfred Michael: "Die Nationalvarietät ‚österreichisches Deutsch' als kakanisches Paradoxon". In: Cwanek-Florek, Ewa/Nöbauer, Irmgard (eds.): *Sprachliche Wechselbeziehungen in der Habsburgermonarchie*. Wien: Poln. Akad. der Wiss. 2015, pp. 11–18.

Gumplowicz, Ludwig: *Der Rassenkampf. Soziologische Untersuchungen. Zweite, durchgesehene und mit Anhang, enthaltend die 1875 erschienene Schrift*

„Rasse und Staat" versehene Auflage. Innsbruck: Verlag der Wagner'schen Univ. Buchhandlung 1909.

Halatschka, Raimund: *Zeitungsdeutsch*. Wien: Pichlers Witwe 1883.

Haugen, Einar: "The analysis of linguistic borrowing". In: *Language* 26, 1950, p. 210–231.

Heinrich, Anton: *Grammatik der neuhochdeutschen Sprache für Mittelschulen und verwandte Anstalten, bes. in mehrsprachigen Ländern, 3. Auflage*. Laibach: Kleinmayr & Bamberg 1875.

Hickey, Raymond: "Internally- and Externally-Motivated Language Change". In: Hernández-Campoy, Juan Manuel/Conde-Silvestre, Juan Camilo (eds.): *The Handbook of Historical Sociolinguistics*. Chichester: Wiley-Blackwell 2012, pp. 387–407.

Hornung, Maria: "Lessiak, Primus (1878‒1937), Germanist". In: *Österreichisches Biographisches Lexikon 1815‒1950*, Bd. 5, Lfg. 22, 1970, pp. 157–156

Hornung, Maria: "Nagl, Johann Willibald (1856‒1918), Germanist und Schriftsteller". In: *Österreichisches Biographisches Lexikon 1815‒1950*, Bd. 7, Lfg. 31, 1976, p. 21.

Hurch, Bernhard: "Schuchardt, Hugo". In: *Neue Deutsche Biographie* 23, 2007, pp. 623–624.

Hurch, Bernhard (eds.): *Hugo Schuchardt Archiv. Postkarte (072-01109)*. Graz 2007 ff. Retrieved at 7.4.2018, from https://schuchardt.uni-graz.at/id/letter/363.

Jungmann, Josef Jakub: *Slownjk česko-německý*. Praha: Knjžecí arcibiskupská knihtiskárna, Josefa wdowa Fetterlowá 1835–1839.

Kim, Agnes et al.: "MiÖ – A framework". In: *DiÖ-Online*. Retrieved at 18. 10. 2018, from https://dioe.at/en/projects/task-cluster-c-contact/mioe-a-framework/.

Kim, Agnes: "Multilingual Lower Austria. Historical sociolinguistic investigations on Wenker's questionnaires". In: Bülow, Lars/Fischer, Ann-Kathrin/Herbert, Kristina: *Dimensionen des sprachlichen Raums: Variation – Mehrsprachigkeit – Konzeptualisierung*. Berlin et. al.: Lang 2019, pp. 187–211.

Kim, Agnes/Newerkla, Stefan Michael: "Das Paradox der Toleranz. Sprachliche Nationalisierung des Mittelschulwesens in Böhmen und Mähren im langen 19. Jahrhundert". In: *Jahrbuch des Bundesinstituts für Kultur und Geschichte der Deutschen im östlichen Europa* 26, 2018, pp. 69–98.

Klaić, Franjo: *Kroatisch-serbischer Dolmetscher. Ein praktisches Handbuch zur schnellen und leichten Erlernung der kroatisch-serbischen Sprache*. Agram: Hartmann 1878.

Koppensteiner, Wolfgang/Kim, Agnes: "Perspectives on change: Language (varieties) contact and language ideologies on German in Austria". In: Lenz, Alexandra N./Maselko, Mateusz: *VARIATIONist (Linguistics) meets CONTACT (Linguistics)*. (forthcoming in 2019).

Krauss, Friedrich Salomon: *Sagen und Märchen der Südslaven in ihrem Verhältnis zu den Sagen und Märchen der übrigen indogermanischen Völkergrupp*en. Leipzig: Friedrich 1883.

Kretschmer, Paul: *Wortgeographie der hochdeutschen Umgangssprache*. Göttingen: Vandenhoeck und Ruprecht 1918.

Lessiak, Primus: "Alpendeutsche und Alpenslawen in ihren sprachlichen Beziehungen". In: *Germanisch-Romanische Monatsschrift* II, 1910, pp. 274–288.

Lebensaft, Elisabeth: "Pötzl, Eduard; Ps. Kleinpetz (1851-1914), Journalist und Schriftsteller". In: *Österreichisches Biographisches Lexikon 1815–1950*, Bd. 8, Lfg. 37, 1980, pp. 149–150.

Mattes, Veronika: "Die Korrespondenz zwischen Ferdinand Blumentritt und Hugo Schuchardt". In: Hurch, Bernhard (ed.): *Hugo Schuchardt Archiv*. Graz 2013. Retrieved at 18. 10. 2018, from http://schuchardt.uni-graz.at/id/person/1160.

Mailhammer, Robert: "Towards a framework of contact etymology". In: Mailhammer, Robert (ed.): *Lexical and structural etymology. Beyond word histories*. Berlin/Boston: De Gruyter Mouton 2013, pp. 9–31.

Matras, Yaron: "Contact, Convergence, and Typology". In: Hickey, Raymond (ed.): *The Handbook of Language Contact*. Chichester: Wiley-Blackwell 2010, pp. 66–85.

Meißner, H.: "Meissner, Johannes Friedrich (1847-1918), Schriftsteller und Journalist." In: *Österreichisches Biographisches Lexikon 1815–1950*, Bd. 6, Lfg. 28, 1974, pp. 201–202.

Moravcsik, Edith: "Universals of language contact". In: Greenberg, Joseph H. (ed.): *Universals of Human Language*. Stanford: Stanford University Press 1978, pp. 94–122.

Muhr, Rudolf: "Grammatische und pragmatische Merkmale des österreichischen Deutsch". In: Muhr, Rudolf/Schrodt, Richard/Wiesinger, Peter (eds.): *Österreichisches Deutsch. Linguistische, sozialpsychologische und sprachpolitische Aspekte einer nationalen Variante des Deutschen. 1. Aufl.* Wien: Hölder-Pichler-Tempsky 1995, pp. 208–234.

Muhr, Rudolf/Schrodt, Richard/Wiesinger, Peter (eds.): *Österreichisches Deutsch. Linguistische, sozialpsychologische und sprachpolitische*

Aspekte einer nationalen Variante des Deutschen. 1. Aufl. Wien: Hölder-Pichler-Tempsky 1995.

Muysken, Pieter: "Halfway between Quechua and Spanish: The case for relexification". In: Highfield, Arnold/Valdman, Albert (eds): *Historicity and Variation in Creole Studies*. Ann Arbor: Karoma 1981, pp. 52–78.

Muysken, Pieter: "Scenarios for Language Contact". In: Hickey, Raymond (ed.): *The Handbook of Language Contact*. Chichester: Wiley-Blackwell 2010, pp. 265–281.

Nagl, [Johann] Willibald: "Die wichtigsten Beziehungen zwischen dem österreichischen und dem čechischen Dialect". In: *Blätter des Vereines für Landeskunde von Niederösterreich, Neue Folge* 21, 1887, pp. 356–388.

Nagl, [Johann] Willibald: "Die wichtigsten Beziehungen zwischen dem österreichischen und dem čechischen Dialect". In: *Blätter des Vereines für Landeskunde von Niederösterreich, Neue Folge* 22, 1888, pp. 417–434.

Newerkla, Stefan Michael: "The seamy side of the Habsburgs' liberal language policy: Intended and factual reality of language use in Plzeň's educational system". In: Rindler Schjerve, Rosita (ed.): *Diglossia and Power: Language Policies and Practice in the 19th Century Habsburg Empire*. Berlin: De Gruyter 2003, pp. 167–195.

Newerkla, Stefan Michael: "Kontaktareale in Mitteleuropa". In: Kątny, Andrzej (ed.): *Słowiańsko-niesłowiańskie kontakty językowe: materiały z międzynarodowej konferencji naukowej zorganizowanej przez Wydział Filologiczny Wszechnicy Mazurskiej Instytut Filologii Germańskiej Uniwersytetu Gdańskiego w dniach 27 - 28 czerwca 2005 r.: Słowiańsko-niesłowiańskie kontakty językowe w perspektywie dia- i synchronicznej. Slawisch-nichtslawische Sprachkontakte in dia- und synchronischer Sicht*. Olecko 2007, pp. 29–48.

Newerkla, Stefan Michael: "Bohemismen (und Slowakismen) in Wien (und Österreich)". In: *tribüne. zeitschrift für sprache und schreibung* 3, 2009, pp. 8–12.

Newerkla, Stefan Michael: "Linguistic Consequences of Slavic Migration to Vienna in the 19th and 20th Centuries". In: Moser, Michael/Polinsky, Maria (eds.): *Slavic Languages in Migration*. Wien, Berlin: LIT 2013a, pp. 247–260.

Newerkla, Stefan Michael: "Vom Nachwirken einer Minderheit: Zuwanderer aus den böhmischen Ländern in Wien". In: *Wiener Slavistisches Jahrbuch. Neue Folge* 1, 2013b, pp. 1–19.

N. N.: "Gumplowicz, Ludwig (1838-1909), Staatsrechtler und Soziologe". In: *Österreichisches Biographisches Lexikon 1815–1950*, Bd. 2. Lfg. 7, 1958, pp. 106–107.

Österreichische Nationalbibliothek: *Austrian Books Online. Public Private Partnership der Österreichischen Nationalbibliothek mit Google. Häufig gestellte Fragen*. 03. 05. 2013. Retrieved at 1. 9. 2018, from https://www.onb.ac.at/fileadmin/user_upload/PDF_Download/ABO_FAQ_de_201304_v1.2.3.pdf.

ÖWB 1951 = *Österreichisches Wörterbuch. Mittlere Ausgabe*. Hrsg. im Auftrag des Bundesministeriums für Unterricht. Wien: Österreichischer Bundesverlag. Jugend & Volk 1951.

ÖWB 1979 = *Österreichisches Wörterbuch*. 35., völlig neu bearbeitete und erweiterte Auflage. Hrsg. im Auftrage des Auftrag des Bundesministeriums für Unterricht und Kunst. Wien: Österreichischer Bundesverlag 1979.

ÖWB 2016 = *Österreichisches Wörterbuch*. Schulausgabe. 43. Auflage. Hrsg. im Auftrag des Bundesministeriums für Bildung. Wien: ÖBV 2016.

Payer, Peter: "Der Chronist der Großstadt. Zum 100. Todestag von Eduard Pötzl (1851–1914)". In: *Wiener Geschichtsblätter*, 3/, 2014, 185–221.

Pohl, Heinz Dieter: "Österreich". In: Goebl, Hans/Nelde, Peter H./Starý, Zdeněk/Wölck, Wolfgang (eds.): *Kontaktlinguistik. Ein internationales Handbuch zeitgenössischer Forschung. 2. Halbband*. Berlin: De Gruyter 1997, pp. 1797–1812.

Republik Österreich, Parlament: *Dr. Harald Ofner*. Retrieved at 4. 4. 2018, from https://www.parlament.gv.at/WWER/PAD_01101/.

Rindler Schjerve, Rosita (ed.): *Diglossia and Power: Language Policies and Practice in the 19ᵗʰ Century Habsburg Empire*. Berlin: De Gruyter 2003.

Ronge, Grete: "Du Bois-Reymond, Emil". In: *Neue Deutsche Biographie* 4, 1959, pp. 146–148.

Schick, Paul: "Kraus, Karl (1874–1936), Schriftsteller". In: *Österreichisches Biographisches Lexikon 1815–1950*, Bd. 4, Lfg. 18, 1968, pp. 230–233.

Schleicher, August: "Über die wechselseitige Einwirkung von Böhmisch und Deutsch". In: *Archiv für das Studium der neueren Sprachen und Literaturen* 9, 1851, pp. 38–42.

Schuchardt, Hugo: *Dem Herrn Franz von Miklosich zum 20. November 1883. Slawo-deutsches und Slawo-italienisches*. Graz: Leuschner und Lubensky 1884.

Schuchardt, Hugo: Zu meiner Schrift 'Slawo-Deutsches und Slawo-Italienisches' (1884 erschienen). In: *Zeitschrift für die österreichischen Gymnasien* XXXVII, 1886, pp. 321–352.

Seidelmann, Erich: "Lautwandel und Systemwandel in der Wiener Stadtmundart. Ein strukturgeschichtlicher Abriss". In: *Zeitschrift für Dialektologie und Linguistik* 2/38, 1971, pp. 145–166.

Silverstein, Michael: "Language Structure and Linguistic Ideology". In: Clyne, Paul R./Hanks, Wiliam F./Hofbauer, Carol L.: *The Elements. A Parasession*

on Linguistic Units and Levels. Chicago, Illinois: Chicago Linguistic Society 1979, pp. 193–247.

Spáčilová, Libuše: "Der gegenseitige Einfluß des Tschechischen und des österreichischen Deutsch in näherer Geschichte und Gegenwart". In: Muhr, Rudolf/Schrodt, Richard/Wiesinger, Peter (eds.): *Österreichisches Deutsch. Linguistische, sozialpsychologische und sprachpolitische Aspekte einer nationalen Variante des Deutschen. 1. Aufl*. Wien: Hölder-Pichler-Tempsky 1995, pp. 327–353.

Steinhauser, Walter: *Slawisches im Wienerischen. 2., verm. und verb. Aufl*. Wien: Verband der wissenschaftlichen Gesellschaften Österreichs 1978.

Steinsberg, Karl Franz Guolfinger Ritter von: *Hanns Klachl, oder: Das Rendevouz in der neuen Allee*. N. a. 1797.

Storfer, Adolf Joseph: "Etwas erinnern – An etwas vergessen". In: *Psychoanalytische Bewegung* IV, 1932, pp. 364–369.

Teweles, Heinrich: *Der Kampf um die Sprache. Linguistische Plaudereien*. Leipzig: Carl Meißner 1884.

Thomason, Sarah G./Kaufman, Terrence: *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press 1988.

Venus, Th.: "Storfer, Adolf Joseph (Albert) (1888–1944), Journalist, Etymologe und Psychoanalytiker". In: *Österreichisches Biographisches Lexikon 1815‒1950*, Bd. 13, Lfg. 61, 2009, pp. 328–329.

Venus, Th.: "Teweles, Heinrich (1856–1927), Journalist, Schriftsteller und Theaterdirektor". In: *Österreichisches Biographisches Lexikon 1815‒1950*, Bd. 14, Lfg. 64, 2014, pp. 273–274.

Watts, Richard J: "Language Myths". In: Hernández-Campoy, Juan Manuel/Conde-Silvestre, Juan Camilo (eds.): *The Handbook of Historical Sociolinguistics*. Chichester: Wiley-Blackwell, pp. 585–606.

Zeman, Dalibor: *Das österreichische Deutsch und die österreichisch-tschechischen Sprachbeziehungen. Ein kulturhistorischer und sprachlicher Abriss*. Dissertation: Universität Wien 2003.

Zeman, Dalibor: *Überlegungen zur deutschen Sprache in Österreich. Linguistische, sprachpolitische und soziolinguistische Aspekte der österreichischen Varietät*. (PHILOLOGIA – Sprachwissenschaftliche Forschungsergebnisse 131). Hamburg: Verlag Dr. Kovač 2009.

Zeman, Dalibor: "Tschechische Einflüsse im Wienerischen auf phonologischer, phraseologischer sowie lexikalischer Ebene. Einige Bemerkungen zu den österreichisch-tschechischen Sprachbeziehungen". In: *brücken ‒ Germanistisches Jahrbuch TSCHECHIEN – SLOWAKEI* 19, 2011, pp. 43–91.

Agnes Kim, Sebastian Scharf, Ivan Šimko

Variation in Case Government of the Equivalent for the Cognitive Verb *to Forget* in German in Austria and Czech

Abstract: This paper investigates the areal variation in case government of the German verb *vergessen* 'to forget' in written standard registers: While the German standard language exclusively recognises constructions with *vergessen* and a direct argument in accusative, the Austrian standard German accepts constructions with a prepositional argument including the preposition *auf* 'on', too. Already since the 19th century scholars have pointed out a similar grammatical variation in case government for the Czech equivalent *zapomínat/zapomenout* 'to forget', considering the situation in Austrian German to reflect Czech influence. Thus, this paper is a first step in the assessment of the plausibility of the language contact explanation for the respective phenomenon. The paper employs corpus linguistic methods to first test the hypothesis that the construction with the prepositional argument is typical for German in Austria. Using contemporary German corpora composed of journalistic texts from Austria, Germany, Switzerland and Liechtenstein, it demonstrates that the construction with the preposition *auf* 'on' occurs considerably more frequently in Austrian texts. Second, the paper evaluates, whether the situation in German in Austria can be attributed to historical language contact. For that purpose, it determines the relations between the two variants of case government and the meaning of the verb in a particular sentence in German in Austria and Czech contrastively. The analysis of corpora of contemporary journalistic texts from Austria and the Czech Republic shows that the constructions with the prepositional object occur considerably more often with a specific meaning of the verb in both languages. For that reason, we conclude that the contact explanation is plausible and requires further research.

Keywords: German in Austria, Slavic, variation, language contact, argument structure, PP-objects, NP-objects

1 Introduction[1]

1.1 Phenomenon and Research Questions

This paper examines the variation in case government of the German cognitive verb *vergessen* 'to forget'. Thus, it investigates one of the most prominent

---

on a long list of linguistic phenomena of German in Austria (Germ. *Deutsch in Österreich*, subsequently *DiÖ*), which have been suspected to result from intense language contact of DiÖ with Slavic languages in general and Czech in particular (cf. Newerkla, this volume and Kim, this volume). In DiÖ, the verb may either govern an argument in the accusative or in the prepositional case with the preposition *auf* 'on' as illustrated by the examples (1) and (2). In German in Germany and Switzerland, on the other hand, only the accusative case (1) is used. In this paper, we will refer to cases such as (1) as the *acc*-construction and to cases represented by (2) as the *prep*-construction.

(1)   Die Heimelf war klar überlegen und spielte sich in einen wahren Rausch,
       vergaß dabei aber nicht das Toreschießen.[2]
       'The home team was clearly superior and played as if in a delirium.
       However, in doing so they did not forget to score goals.'[3]
       (DeReKo-2015-II/Burgenländische Volkszeitung, 05.08.2010)

---

2   In the examples, the following syntactical elements are highlighted throughout the text: The verb in focus, i.e. the equivalents of 'to forget' are typeset in **bold characters**, the direct argument is underlined and the preposition in focus, either part of the direct argument and thus of the *prep*-construction, or of an adverbial, in *italics*.

3   Unless otherwise stated, the examples were translated by the authors.

(2)   Nach dem schnellen Tor haben wir <u>aufs</u> Tore-Schießen **vergessen**.
       'After the quick goal we forgot to score.'
       (DeReKo-2015-II/Burgenländische Volkszeitung, 18.03.2010)

Historically, a second *prep*-construction, a prepositional argument with the preposition *an* 'at' is attested in newspapers from various crown lands of the Habsburg monarchy (3). Both *prep*-constructions have repeatedly been treated as contact phenomena (cf. Kim, this volume). However, since it is neither known nor used in contemporary DiÖ, we are not concerned with it in the present analysis. Unless explicitly stated, the term *prep*-construction thus refers to examples such as (2) and not (3) in the context of this paper.

(3)   Man war verblüfft und **vergaß** <u>an</u> die heitere Bewegung und den
       witzigen Dialog, […].
       'Everybody was amazed and forgot about the cheerful motion and the
       witty dialogue,
       […].'
       (ANNO/*Blätter für Musik, Theater und Kunst*, 12.03.1872).

For the Czech equivalent *zapomínat* (ipf.)/*zapomenout* (pf.) 'to forget' both the *acc*-construction (4) and the *prep*-construction (5) are common. Therefore, the phenomenon in DiÖ has widely been proposed to result from language contact with Czech from the late 19th century (e.g. Teweles 1884, p. 11[4]) up until recently (e.g Muhr 1995, p. 226, Zeman 2003, pp. 275–277, Newerkla 2007, p. 280, Blahak 2015, p. 509).

(4)   **Zapomíná** <u>text</u> a svým vzhledem pobuřuje okolí.
       'He forgets the text and incites the surroundings with his looks.'
       (SYN2015/Rytmus života, 30/2010)

(5)   Je fakt, že se najednou všichni hrnou do útoku a **zapomínají** <u>na</u> obranu.
       'It is a fact that all of a sudden, everybody hurls oneself to attack and
       forgets about the defence.'
       (SYN2015/Sport, 19.01.2010)

---

4   Note, that Teweles (1884) does not refer to what has been defined as DiÖ above, but rather to German in Prague. However, since Teweles' contemporaries such as Schuchardt (1884) describe such local varieties and contact situations to have an impact on German all over the Habsburg monarchy, which later on considerably shaped DiÖ, Teweles' remark is a valuable piece in the puzzle.

Interestingly, none of the aforementioned authors provides a valid scientific study either on the phenomenon in focus or on related phenomena, which may prove or reject the language contact explanation. Eventually, Blahak (2015, p. 509) concludes that from a contemporary perspective it cannot be decided anymore, whether the variation in case government of *vergessen* 'to forget' actually can be traced back to language contact with Czech.

This paper challenges this view and seeks to investigate whether the contact explanation for the variation in case government of Germ. *vergessen* 'to forget' is feasible. We presuppose that the historical language contact scenario in the area, which is nowadays the Czech Republic and eastern Austria, facilitated language change on the syntactic level and that from a historical sociolinguistic perspective the contact explanation is thus plausible (cf. e.g. Kim 2019 for Lower Austria).

We approach the phenomenon in focus synchronically in a corpus linguistic and contrastive framework as follows: First, we test the hypothesis, that the *prep*-construction is indeed restricted to DiÖ. This is important in order to exclude a broader spread of the *prep*-construction in spoken and written registers of German. However, even a restriction of the *prep*-construction to Slavic-German contact areas would support the plausibility of the contact explanation. In a second step, we verify the second hypothesis, i.e. that the *prep*-construction in DiÖ and Czech is semantically and syntactically identical or at least similar. Thus, we investigate, whether the patterns of case variation of the respective equivalents of English *to forget* actually correspond in DiÖ and Czech. Semantic and morphosyntactic features are taken into account to identify those patterns. Both hypotheses are expected to hold true. This would indicate that the contact explanation is plausible from a contrastive perspective and that it should thus be further investigated, e.g. diachronically.

Before turning to the outlined research questions in section 2, 3 and 4, we review seminal historical and contemporary dictionaries of German and Czech with regard to the codification and standardisation of the *prep*-construction. This information will be crucial when it comes to the interpretation of our results. Then, we turn to the contact explanation as such and give a brief overview of its development and tradition.

1.2 Codification and Standardisation Issues

In this section, seminal dictionaries of Czech and German from the 19th century and selected ones with considerable impact from the 20th century are reviewed in order to assess the codification or the degree of standardisation of the *prep*-construction. The focus primarily lies on whether the dictionaries cover the

prep-construction and which information they give regarding its regional or stylistic distribution. The semantics will be taken into account in section 3.2.

The first benchmark in German lexicography (cf. Kühn/Püschel 1990, p. 2058), the dictionary by Adelung (1st edition: 1774–1786, 2nd edition: 1793–1801), includes the *prep*-construction next to the primary *acc*-construction and a *gen*-construction, i.e. cases in which *vergessen* governs an argument in genitive case. It characterises the *prep*-construction as "common in Upper German but completely inacceptable in High German"[5] (Adelung 1801, col. 1045), i.e. inacceptable in the literary language.[6]

The German Dictionary by the Brothers Grimm (*Deutsches Wörterbuch* = DWB 1854–1961[7]) cites Adelung and explicitly adds that the *prep*-construction is used "scarcely and mainly in dialects"[8] (DWB 25, col. 421). Besides the examples given by Adelung (1801), the German Dictionary adds two further examples that it classifies as "Upper Austrian".

Thus, the main codifying German dictionaries from the 19th century include the *prep*-construction, but assess it as a regional, Upper German variant, which does not belong to the literary, the standard register. Similarly, contemporary dictionaries of German record the *prep*-construction with *vergessen* 'to forget', but mark it as "regional, (especially) Austrian, colloquial"[9] (Dictionary of Contemporary German, *Wörterbuch der deutschen Gegenwartssprache* = WDG 1964–1977, cited according to DWDS 2017) or as "south German, Austrian"[10] (Dudenredaktion 2017).

---

5   Original: "Hingegen die Ausdrückung der Sache mit dem Vorworte auf, welche gleichfalls im Oberdeutschen üblich ist, ist im Hochdeutschen völlig ungangbar."

6   To interpret this description adequately, one needs to be aware of the underlying definitions of the terms Upper German (*Oberdeutsch*) and High German (*Hochdeutsch*). These cannot be interpreted against a modern dialectological background, in which High German is used to describe all dialects that have undergone the so-called High German consonant shift. Upper German is then defined to be the southern subgroup of High German dialects with Middle German (*Mitteldeutsch*) being the other, northern sub-group. Adelung's notion of High German rather coincides with its widespread general use referring to the standard register of German.

7   The part of volume 25 that contains the lemma *vergessen* 'to forget' was published in 1889.

8   Original: "*selten und hauptsächlich nur in den mundarten wird das object nicht im accusativ, sondern durch eine präposition verbunden beigefügt:* auf etwas vergessen, ,*im oberdeutschen üblich, im hochdeutschen völlig ungangbar*'"

9   Original: "landschaftlich, (besonders) österreichisch, umgangssprachlich"

10 Original: "süddeutsch, österreichisch"

Both editions of the Dictionary of German Standard Varieties (*Variantenwörterbuch des Deutschen*, Ammon et al. 1st edition: 2004, p. 829, 2nd edition 2016, p. 780) assign the *prep*-construction to the Austrian standard variety. The Austrian Dictionary (*Österreichisches Wörterbuch* = ÖWB 43rd edition: 2016, p. 772) adds the preposition *auf* 'on' in square brackets in usages such as in example (6), thus indicating that it is acceptable.

(6)   [*auf*] den Geburtstag **v.**
        'to forget [about] the birthday'
        (ÖWB 2016, p. 772)

Among the Czech dictionaries of the 19th century, the Czech-German Dictionary (*Slovník česko-německý*) by Jungmann (1835–1839) had a codification impact similar to Adelung (1793–1801). According to Jedlička (1990, p. 2280), its main objective was to emancipate Czech concerning its societal-communicative capacity and thus develop it towards a contemporary literary language. Jungmann (1835–1839) includes the *acc*-, *gen*- and the *prep*-construction equally and does not distinguish between them with regard to style, register or areal distribution. Interestingly, the information on argument structure of the verb *zapomenout* 'to forget' as given in (7a) is translated to German with both the *acc*- and the *prep*-construction (7b).

(7)   a.   **zapomenouti** něco, něčeho, *na* něco
              'to forget something, about something'
              (Jungmann 1839, p. 525)

        b.   etwas, *auf* etwas **vergessen**
              'to forget something, about something'
              (Jungmann 1839, p. 525)

The main standardising dictionary of Czech, the Reference Dictionary of the Czech language (*Příruční slovník jazyka českého* = PSJČ 1935–1957), also does not differentiate between the *acc*- and *prep*-construction with regard to style, register or areal distribution. The *gen*-construction, on the other hand, is marked as typical for the literary (č. *knižní*) style. A later dictionary of Czech, the Dictionary of Written Czech (*Slovník spisovného jazyka českého* = SSJČ 1960–1971) basically follows the PSJČ but classifies the *gen*-construction not only as literary, but also as outdated (č. *zastaralý*). Hence, both the *acc*- and the *prep*-construction with *zapomínat/zapomenout* 'to forget' are part of the standard written register of Czech.

From reviewing the dictionaries, we conclude that in contrast to Czech, the *prep*-construction in German is not part of a supra-regional standard register. It is mostly assigned to colloquial or dialectal, i.e. spoken registers of DiÖ, but also acceptable in the spoken and written Austrian standard register. Hence, we presuppose that in written texts and, consequently, in newspaper articles the *prep*-construction is underrepresented in comparison to its use in colloquial, spoken language.[11] This information will prove vital for the interpretation of our results.

1.3 The Language Contact Explanation[12]

An early interpretation of the *prep*-construction as a Czech influence on DiÖ dates back to Teweles (1884). From the involved, hardly neutral perspective of a German from Prague, the journalist satirically describes the administrative register of German in Bohemia and explains the widespread use of *prep*-constructions with either the preposition *auf* 'on' or *an* 'at' as an effect of bilingualism, i.e. of those people "who claim to master both languages, German and Czech, but actually only speak Czech German or germanised Czech"[13] (Teweles 1884, p. 11).

In more recent times, the proposition about Czech influence has been reiterated by Muhr (1995, p. 226) and Zeman (2003, p. 275 ff.), who both consider this construction typical for the standard register of DiÖ. Muhr (1995) does not cite any thorough studies on the phenomenon, nor does he provide one himself. He simply gives the fairly brief information that the *prep*-construction results from language contact with Czech and Slovak in brackets and, thus, accepts it as a

---

11 Currently, we are not aware of a valid study that investigates the distribution of the *prep*-construction in spoken registers of DiÖ. This paper cannot close this gap as it focuses on the *prep*-construction with *vergessen* 'to forget' in written texts, i.e. newspapers.

12 This section gives a brief overview of explanations specific for the construction in focus. For a more detailed contextualisation see Kim, this volume.

13 "Denn daß ein solches Bedürfniß, einige schmählich verstümmelte und verwandelte Laute als deutsche Sprache auszugeben, thatsächlich besteht, dafür ist doch der Umstand Beweises genug, daß sich bei uns so viele Leute mit der Kenntniß beider Landessprachen ausweisen, die offenbar nur ein čechisches Deutsch und ein germanisirtes Čechisch sprechen, Leute, welche an dem kunstvollen Gebäude unseres Amtstyles bauen und denen wir die Nachdemisirung der deutschen Sprache verdanken. Das Schlimmste daran ist, daß diese verdorbene Sprache uns allgemacht ganz vertraut klingt, daß wir oft darauf oder daran vergessen (das ist auch eine jener čechisierten Wortfügungen), daß diese oder jene Wendung, dieses oder jenes wort gar nicht deutschen Ursprungs ist und nur ein örtliches Verständnis findet." (Teweles 1884, pp. 10–11).

fact. Similarly, when it comes to assessing the phenomenon, Zeman (2003) only cites consultations with different Austrian linguists, who either argue for a "general influence of the monarchy" or think that the contact explanation should be treated with caution. However, Zeman (2003) attempts to empirically approach the *prep*-construction with *vergessen* 'to forget' amongst other alleged contact phenomena, that display variation with regard to prepositions. In a questionnaire distributed in Vienna and Lower Austria, he asks the respondents to fill in a preposition in a gap text, which provides a single context per phenomenon, e.g. (8).

(8)   Er **vergaß** völlig _____ das Abendessen.
        'He completely forgot _____ dinner.'
        (Zeman 2003, p. 357)

Unfortunately, the author did not consider that in (8), the preposition does not alternate with another preposition, but with a direct argument in accusative, i.e. the *acc*-construction. Both the word order given in (8) and the exercise trigger the *prep*-construction. Therefore, it is not surprising that the *prep*-construction with *vergessen* 'to forget' turns out to be the most frequently used one out of Zeman's sample – depending on their age and origin, between 91 and 100 % of the respondents seem to actively use it. However, the study does not provide reliable information on the frequency of the *prep*-construction in comparison to the *acc*-construction either. The only information that can be deduced from Zeman's study is that the *prep*-construction is (still) known in German in eastern Austria. Newerkla (2007, p. 280 f.) focuses on eastern Austrian (or Viennese) registers, too, and explains the phenomenon as a result of the immigration of Czechs to Vienna.

Not all scholars adhered or adhere to the explanation of the phenomenon as Czech influence on German. Kretschmer (1918, p. 7) considers the *prep*-construction with *auf* 'on' a native construction, while he assesses the variant with the preposition *an* 'at' as "Jewish" influence. However, he agrees with other sources that both *prep*-constructions with *auf* 'on' and *an* 'at' are limited to southern German dialects. The framing of the *prep*-construction with *an* 'at' as Jewish reoccurs in Viennese journalistic discourse in the early 20th century, especially in texts by Karl Kraus (cf. Kim, this volume).

From reviewing several of these (contact) explanations, Blahak (2015, p. 509) concludes – as already cited above – that from a contemporary perspective the origin of the *prep*-construction in DiÖ cannot be explained anymore. The present paper does not allow any judgment about the origin of the construction either. The question, whether it is a Czech or Slovak innovation which spread

to DiÖ, whether it was adopted by all of these languages from a third language, or if it is a phenomenon of areal convergence requires further investigation in follow-up studies.

In this context, it should be noted that according to a brief review of contemporary corpus data,[14] the *prep*-construction with *na* 'on, at' is not a common construction across all Slavic languages. All Slavic languages know the variation of an *acc*-construction and a *prep*-construction as well, but some use other prepositions in the *prep*-construction, e.g. Pol. *o* 'about', and in the East Slavic languages: Russ. *o* or *pro* 'about', Ukr. *pro* 'about'. In Belorussian both *pra* 'about' or *na* 'on, at' are possible. The West Slavic languages except for the aforementioned Polish, i.e. the Sorbian languages, Czech and Slovak use the *prep*-construction with *na* 'on, at' very frequently. In the South Slavic languages the frequency of the *prep*-construction with *na* 'on, at' seems to decrease on the north-south axis: In Slovenian, many examples can be readily found. In Croatian and Serbian it is also attested, even in Macedonian it seems to appear in restricted environments. In Macedonian and Bulgarian the equivalent of Engl. 'to forget' may govern a *prep*-construction with *za* 'for'. Hence, a core area of the *prep*-construction with *na* 'on, at' in the Slavic languages can probably be identified in Central Europe, most likely in those areas that were part of the Habsburg monarchy.

As argued by Newerkla (this volume), the role of Yiddish in the context of language contact areal convergence phenomena in Central Europe must not be neglected. The only dictionary of Yiddish, that gives information on the valency of the verb *fargesn* 'to forget', lists a *prep*-construction *fargesn af* as the equivalent of 'forget about' (cf. Schaechter-Viswanath/Glasser 2016, p. 490). The etymologic connection of Germ. *auf* and Yiddish *af* is obvious. Hence, the description of the German *prep*-construction with *an* 'at' as "Jewish" as propagated, e.g. by Kretschmer (1918) requires reconsideration.[15]

---

14 Our observations are based on brief analyses of corpus data from the Slavic *Araneum*-webcorpora (cf. Benko 2014) and the subcorpora of Slavic languages in the InterCorp (cf. Rosen/Vavřín/Zasina 2017). Scharf (2018) investigates the equivalent patterns of case variation in Polish, Russian and Ukrainian. Detailed analyses for the other languages as well as a comparative analysis remain to be conducted.

15 In the Corpus of Modern Yiddish (cf. CMY), examples of a *prep*-construction with the preposition *vegn* 'about, regarding, on' are attested, too. Hence, such a reconsideration should aim at describing factors of variation in case government in contemporary and historical Yiddish and consider several contact languages from German, Russian and Hebrew to English.

2 Areal Distribution of the *Prep*-Construction in German

As shown in chapter 1.2., seminal German historical dictionaries (Adelung 1793–1801, DWB 1854–1961) mention the use of the *prep*-construction as regionally restricted and dialectal, and not being part of the written register of German. Contemporary dictionaries mark the construction as regional and colloquial with a core area in Austria. This chapter provides detailed information on the distribution of the *prep*-construction in regional colloquial and dialectal registers, but also in written registers of German. Thus, it evaluates the supposed areal restriction of the *prep*-construction within the German language area.

2.1 The Distribution in Regional Colloquial Registers of German

The *prep*-construction is subject to a few publications dealing with the geographical distribution of German colloquial registers. According to Kretschmer (1918), the *prep*-construction is part of a German colloquial standard and mostly used in Bavaria and Austria. It is, however, also attested for Baden and Zweibrücken[16] (cf. Kretschmer 1918, p. 7). Zehetner (2014) is particularly dedicated to the historical region of Old Bavaria (*Altbayern*), which is de facto identical to the former Electorate of Bavaria. According to this dictionary, the *prep*-construction occurs in this region with two possible meanings: a) 'to not think about something on time', cf. (9a), and b) 'to not care about something or somebody anymore', cf. (9b).

(9)   a.   Vergiss nicht *drauf*, dass du den Brief aufgibst!
               'Don't forget to post the letter!'
               (Zehetner 2014, p. 362)

        b.   Hast *auf* deine alte Mutter ganz v[ergessen]?
               'Did you completely forget about your old mother?'
               (Zehetner 2014, p. 362)

Another approach can be found in the Word Atlas of the German Colloquial Language (*Wortatlas der deutschen Umgangssprache,* Eichhoff 1993). Eichhoff's assumptions about the areal distribution of the construction are based on questionnaires and show that the *prep*-construction is not as widespread in

---

16 Zweibrücken is a town in the central-western state of Rhineland-Palatinate. Baden is part of the south-western German state Baden-Württemberg, Kretschmer lists the towns of Rastatt, Karlsruhe, Freiburg and Donaueschingen, i.e. towns from south-western parts of Baden (Kretschmer 1918, p. 7).

Southern Germany as the dictionaries above might suggest, but that it is predominantly used in colloquial DiÖ (cf. Eichhoff 1993, p. 35).

2.2 The Distribution in Dialectal Registers of German

The following section focuses on German dialects and reviews information on the occurrence or rather records of the *prep*-construction from the so-called Major Landscape Dictionaries of German dialects (*Großlandschaftliche Wörterbücher der deutschen Dialekte*) listed in Friebertshäuser (1983) and König et al. (2015).[17]

In the majority of the Major Landscape Dictionaries, there is no record of the *prep*-construction for the verb *vergessen* 'to forget'.[18] Those dictionaries, which mention the *prep*-construction, also list the *acc*-construction. This regards Bohemian, Moravian and Silesian (cf. Engels et. al. 1982–2010, [10]) as well as Upper Saxon dialects (cf. Bergmann/Helm 1994–2003, [11]). In the case of Upper Saxon dialects, the *prep*-construction is restricted to the regions of Upper-, Eastern- and Western Lusatia (*Ober-, Ost- und Westlausitz*), where the Sorbian languages are spoken, as well as (rarely) the Ore Mountains (*Erzgebirge*). Both regions are adjacent to or overlapping with Slavic language areas.

(10)   *aff*s Böia(r) hams wuhl vagess'n
         'they probably completely forgot the beer'
         (Engels et al. 1982–2010, p. 138)

---

17 The dictionaries subsumed under this category cover the whole contemporary as well as historical German language area. Generally, these dictionaries do not only aim at exclusively dialectal lexemes but also target forms of the standard variety in their special, dialectal meaning (cf. Friebertshäuser 1982, p. 1283). This is the case for the German verb *vergessen* 'to forget': Although being part of the standard register semantical and grammatical properties may differ in certain dialect areas. However, some of these dictionaries are not finished yet. In some cases, unpublished material could be obtained from the dictionary projects. Several unfinished dictionaries mentioned by Frieberthäuser (1982) made progress during the last decades. A more current overview can be found in König et al. (2015).

18 In a few dictionaries, there is no entry for the verb at all. In others, either the corresponding volume is not finished yet or the verb is simply not registered. This concerns the regions and dialects of the Rhineland, Westphalia, Eastphalia (*Ostfalen*), the Transylvanian Saxons (*Siebenbürgen Sachsen*), Pomerania, Switzerland, Luxembourg, Prussia and Franconia. A rather isolated but interesting *prep*-construction with the preposition *um* 'around' is recorded for Low German in Schleswig-Holstein.

(11)   a hutte ganz <u>dr*uff*</u> **vergassen**
        'and he had completely forgotten about it'
        (Bergmann/Helm 1994–2003, p. 461)

Unfortunately, the respective dictionaries of Bavarian dialects spoken in Bavaria (*Bayerisches Wörterbuch*, BWB) and Austria (*Wörterbuch der bairischen Dialekte in Österreich*, WBÖ) have not yet been published. Nevertheless, unpublished material from these dictionary projects shows the use of the *prep*-construction, e.g. (12) for Bavaria:[19]

(12)   da muass ma *auf*'s Essen und 's Trinken vagessn
        'one is bound to forget about food and drinks'
        (Material from the Bavarian Dictionary, BWB)

Among the material obtained from the Austrian Academy of Science a single example of the *prep*-construction can be found. Example (13) was recorded in Egerland, a region in the far north west of Bohemia that has already been excluded from the focus area of the WBÖ.[20]

(13)   I ho hali dr*af* vagessa
        'I had simply forgotten about it'
        (Material from the Dictionary of Bavarian dialects in Austria, WBÖ)

Furthermore, the *prep*-construction is recorded for Mainz, where it is attributed to a former Austrian garrison in the town (cf. Maurer/Mulch 1965–2010, p. 492). On the other hand, the use of the *prep*-construction in some regions of Baden as mentioned by Kretschmer (1918, p. 7) is not confirmed in the corresponding dictionary (cf. Ochs/Müller 1925–2009, p. 60).

To sum up, dialectological dictionaries register the *prep*-construction for dialects of East-Middle German and East-Upper German (Bavarian). Apart from that, there is no evidence for the *prep*-construction with the preposition *auf* 'on' in German dialects. Concerning colloquial German, contemporary publications support the assumed restriction to DiÖ and to a smaller degree to some regions of Bavaria.

---

19 We thank Philipp Stöckle, current editor of the WBÖ for generously retrieving and providing the according information from the main catalogue of the WBÖ. We also thank Anthony Rowley for providing us with material from the BWB.

20 Unfortunately, there is no record of the *prep*-construction from the area of present-day Austria among the material.

2.3 Questionnaire Surveys on the Spread in DiÖ

As already mentioned we are not aware of an empirical survey, which seeks to describe the areal distribution of the *prep*-construction in colloquial and dialectal registers of DiÖ. With regard to the standard register of DiÖ, however, and besides Zeman's (2003) methodically flawed attempt, two recent self-report questionnaire surveys (cf. Wiesinger 2015, Břenek 2017) investigate the current use of the *prep*-construction among Austrian students.[21] As Břenek (2017) replicates the questionnaire used by Wiesinger (2015), both studies collected their data on the *prep*-construction with the exact same gap sentence (14).

(14)   Ich habe leider _____ Treffen **vergessen**.
        'I unfortunately forgot _____ meeting.'
        (Wiesinger 2015, p. 114; Břenek 2017, p. 62)

This sentence does not syntactically trigger the *prep*-construction, as the stimuli in Zeman (2003) does. However, the survey design still presupposes that the *acc*- and *prep*-construction are synonymous and do not vary along syntactic or semantic lines, but only diatopically and diastratically. Therefore, the results require cautious reconsideration, even though they give interesting insight into the acceptability of the *prep*-construction as part of the standard variety of DiÖ.

Wiesinger (2015, p. 93) explicitly requested the respondents to report the variant, which they would use in written texts. Břenek (2017 [appendix]) on the other hand only asked for the most frequently used variant. Table 1 gives an overview of the results from the states of Austria, in which Bavarian dialects are spoken, as reported by the respective authors.

With regard to the phenomenon in focus, the two surveys significantly diverge: Břenek's (2017) results suggest a widespread use of the *prep*-construction throughout the investigated parts of Austria, whereas in Wiesinger's (2015) sample, the *prep*-construction seems to be more acceptable in the standard register of DiÖ for students from eastern Austria. These differences are probably related to Wiesinger's (2015) explicit interest in the written standard register. He

---

21 Wiesinger (2015) conducted his research in 2012 and 2013 at the universities of Vienna, Graz, Klagenfurt, Salzburg and Innsbruck. All his 395 informants were students of German philology (cf. Wiesinger 2015, pp. 93–94). Břenek (2017) used an online questionnaire and collected responses in 2014. Even though he did not restrict his pool of informants to university students, but rather intended to study the "younger and higher educated population of Austria", 234 out of 300 respondents are students at several Austrian univiersites, but not necessarily students of German philology (cf. Břenek 2017, pp. 53–54, pp. 71–72).

Tab. 1: Overview of Wiesinger's (2015) and Břenek's (2017) results

state	Wiesinger 2015: 115		Břenek 2017[a]	
	acc-constr.	*prep*-constr.	*acc*-constr.	*prep*-constr.
Vienna	80.3 % (53)	**19.7 % (13)**	50 %	**50 %**
Burgenland	85.7 % (12)	14.3 % (2)	50 %	**50 %**
Lower Austria	79.6 % (43)	**20.4 % (11)**	70 %	30 %
Upper Austria	92.0 % (46)	8.0 % (4)	65 %	35 %
Styria	87.5 % (42)	12.5 % (6)	75 %	25 %
Carinthia	84.1 % (37)	15.9 % (7)	80 %	20 %
Salzburg	88.6 % (39)	11.4 % (5)	50 %	**50 %**
Tyrol	94.4 % (34)	5.6 % (29)	/	/
	87.0 % (342)	**13.0 % (51)**	**68 %**	**32 %**

[a]Note, that Břenek (2017) does not report his results in numbers, but only in diagrams, which compare the percentages of each variant for two states in his appendix. Table 1 gives the approximate percentages and subsumes the results for *das* 'the (ACC)', *unser* 'our (ACC)', *ein* 'a (ACC)' under the category *acc*-construction. Břenek (2017) does not report results for Tyrol at all due to an insufficient number of respondents. The exact number for the overall usage is given in the author's comparison with Wiesinger's (2015) results (cf. Břenek 2017, p. 90).

assumes the low acceptance of the *prep*-construction in written texts in states such as Upper Austria and Tyrol to coincide with the widespread dialect use in these states and hold school education responsible for the underrepresentation of the feature in written texts. According to him, several students report a discrepancy of their use of the *prep*-construction in spoken and written language, which is facilitated by the school teachers preferring the *acc*-construction over the "Austrian" *prep*-construction (cf. Wiesinger 2015, p. 115).

However, these questionnaire surveys indicate that young and educated speakers of DiÖ accept and actively use the *prep*-construction in the spoken standard register and – to a certain degree – in the written standard register of DiÖ.

2.4 Evidence from Contemporary Corpora

In order to confirm the restriction of the *prep*-construction to a written standard register of DiÖ in comparison to other standard registers of German, evidence from the German Reference Corpus (*Deutsches Referenzkorpus*: DeReKo-2015-II) created by the Institute for the German Language (*Institut für Deutsche Sprache*, IDS) is collected. This corpus contains texts from all German-speaking

countries, i.e. Germany, Austria, Switzerland and Luxembourg, and its web application COSMAS II allows the creation of sub corpora.[22]

For our purposes, comparable sources from the four German-speaking countries are required. In the case of Luxembourg, the genre of journalistic texts is the only one which is sufficiently represented. Additionally, journalistic texts are available to a larger degree and can easily be located geographically. Hence, our sub corpus consists of a collection of national and regional newspapers: four from Germany, three from Austria, two from Switzerland and one from Luxembourg from the period 2010 to 2015.[23]

A general shortcoming of the DeReKo is the lack of morphologically or syntactically annotated corpora of suitable size for several countries. Hence, our sub corpus is merely lemmatised. Therefore, a search query that aims at the verb *vergessen* 'to forget' governing a prepositional argument with *auf* 'on' in several syntactic configurations has to exclude contexts, in which the preposition is part of an adverbial phrase as illustrated by example (15).

(15)   […] dass sie <u>eine Pfanne mit Fett</u> *auf* dem […] E-Herd **vergessen** hatte.
         '[…] that she had forgotten a pan with fat on the stove.'
         (DeReKo-2015-II/Burgenländische Volkszeitung, 13. 01. 2010)

It is crucial to keep in mind that in doing so, a possible exclusion of suitable examples for the *prep*-construction in focus was accepted; simultaneously, it was acknowledged that some examples may not show the construction.[24] The request aims to measure the relative frequency of the *prep*-construction in the single sources. In interpreting the results given in Tab. 2 and 3, the second column (relative frequency) must be taken into account. It shows the relatively higher frequency of the *prep*-construction in data from Austria.

---

22 Accessible online: http://www.ids-mannheim.de/cosmas2/web-app/ (29.03.2018).

23 Our sub corpus consists of the following newspapers: Germany: *Hamburger Morgenpost, Hannoversche Allgemeine, Nürnberger Zeitung, Süddeutsche Zeitung;* Austria: *Burgenländische Volkszeitung, Niederösterreichische Nachrichten, Die Presse;* Switzerland: *Die Südostschweiz, St. Galler Tagblatt;* Luxembourg: *Luxemburger Tageblatt.* We are aware of the fact that we are biased regarding the diversity of text types. As the availability of a larger amount of texts from several German speaking countries plays the most important role, these shortcomings of the corpus are overlooked.

24 According to this, the search query is not as self-explanatory as expected. However, it was verified in several requests and shows a very high hit rate.
    ((&vergessen %w0 vergessene+) /w1:6,s0 (auf+ %w0:1 (der oder +einer oder dieser oder dem oder +einem oder diesem oder,))) %w0 "und" oder "oder"

Tab. 2: Relative frequency of the *prep*-construction in newspapers

hits	relative frequency	texts	from	until	source	country
937	**6.847 pMW**	923	2010	2015	Niederösterreichische Nachrichten	**A**
172	**6.607 pMW**	169	2010	2015	Burgenländische Volkszeitung	**A**
430	**5.339 pMW**	414	2010	2015	Die Presse	**A**
236	1.803 pMW	2017	2010	2015	Süddeutsche Zeitung	D-S
44	1.563 pMW	42	2010	2015	Luxemburger Tagblatt	L
48	1.341 pMW	48	2010	2015	Hamburger Morgenpost	D-N
71	1.307 pMW	67	2010	2015	Nürnberger Zeitung	D-S
81	1.110 pMW	75	2010	2015	Die Südostschweiz	CH
137	0.985 pMW	130	2010	2015	St. Galler Tagblatt	CH
8	0.739 pMW	8	2010	2014	Hannoversche Allgemeine	D-N

Tab. 3: Relative frequency of the *prep*-construction according to countries

hits	relative frequency	texts	from	until	country
1539	**6.323 pMW**	1506	2010	2015	**A**
363	1.566 pMW	340	2010	2015	D
44	1.536 pMW	42	2010	2015	L
218	1.028 pMW	205	2010	2015	CH

The total number of hits is 2164. Fig. 1 shows the results according to the source. A significantly higher relative frequency is attested for all sources originating from Austria, i.e. in the *Burgenländische Volkszeitung*, *Niederösterreichische Nachrichten* and *Die Presse*. Tab. 3 presents the same results according to countries and maintains the impression from Tab. 2.

These results confirm the hypothesis, that the *prep*-construction with the verb *vergessen* 'to forget' occurs significantly more frequently in written DiÖ than in other written registers of German. Additionally, they indicate that the semi-standardised *prep*-construction at least occasionally appears in newspapers and magazines in Austria, which makes this genre an appropriate one for the closer examination of the construction.

3 Contrastive Analysis of Czech and DiÖ: Corpora and Annotation

3.1 Research Question, Data and Caveats

This chapter is dedicated to the second, contrastive step in the analysis: It identifies patterns of case variation of the respective equivalents of English *to forget* in contemporary DiÖ and Czech and examines whether they actually correspond semantically and/or (morpho-)syntactically. Again, the data is obtained from corpora of written, journalistic texts.

For both languages, the searched corpora were meant to be as comparable as possible with regard to text size and the year of publication. According to these criteria, subcorpora were created from the reference corpora of German and Czech (cf. Tab. 4). Both corpora were searched for all occurrences of the language-specific lemmas of interest, namely the lemma Germ. *vergessen* 'to forget', and for both the imperfective and perfective lemmas Cz. *zapomínat* and *zapomenout* 'to forget', which resulted in the number of hits also shown in Tab. 3.[25] From those, every fourth concordance was manually extracted and annotated according to the schema described in chapter 3.2 and 3.3 (= annotated hits). For the analysis of both Czech and DiÖ, only those examples were taken into account, which display a construction with a direct argument, i.e. the *acc*-, *prep*- and *gen*-construction (= analysed hits). Consequently, reflexive constructions, constructions with a direct and indirect argument as well as occurrences of the respective verb without any arguments were eliminated from the sample.

The initial proposition that the given selection criteria would lead to sufficiently comparable corpora did not prove entirely correct. Whereas the DiÖ corpus comprises only four different sources, all of which are weekly newspapers and news magazines, the Czech corpus contains various sources: The 240 hits analysed in this paper stem from all together 70 different newspapers and magazines. Taking the corpus size and the number of total hits into account, it seems likely that the Czech equivalent occurs significantly more often in the Czech corpus than in the German. To test this hypothesis, the normalised frequency of the respective lemmas in the DiÖ and Czech corpus was compared with the Log-likelihood calculator provided by the University of Lancaster.[26] This tool helps to compare relative frequencies of hits in different corpora effectively

---

25 A preliminary version of this paper, which was presented as a poster in April 2017 analysed the complete set for each language (cf. Kim/Scharf 2017).

26 Cf. http://ucrel.lancs.ac.uk/llwizard.html, 09. 03. 2018.

Tab. 4: Corpora and hits for the equivalents of 'to forget' (overview)

	DiÖ	*Czech*
source	DeReKo-2015-II, archive "W" (IDS Mannheim, cf. IDS 2015)	SYN2015 (Czech National Corpus, cf. Křen et al. 2015; 2016)
language (area)	DiÖ (Austria)	Czech
publication year	2010	2010
genre	four selected newspapers and magazines[a]	all newspapers/magazines available
size	11 217 843 tokens	7 981 747 tokens
lemmatisation	yes	yes
total hits	999	961
annotated hits	249	240
analysed hits	232	236

[a]The four selected newspapers are: *Burgenländische Volkszeitung*, *NEWS*, *Falter* and *profil*. A closer discussion of these sources can be found in chapter 4.3.

(cf. Rayson/Garside 2000). This analysis uses the total number of hits and not the annotated or analysed number. The results show that *vergessen* 'to forget' is 26.11 % less likely to occur in the DiÖ corpus, than *zapomínat/zapomenout* 'to forget' in the Czech corpus.[27] Hence, the aforementioned hypothesis proves correct.

Additionally, it has to be taken into account that the chosen Czech subcorpus reflects the composition of the underlying SYN2015 and thus comprises approximately 60 % (4 751 614 tokens)[28] so-called traditional journalism and 40 % (3 230 133 tokens) leisure magazines[29] (cf. Cvrček/Richterová 2018). The German

---

27 *%DIFF*: –26.11, *LL*: 44.62; The result is highly significant, because according to the developers the *LL* (Log-likelihood) must be above 3.84 for the difference to be significant at the *p* < 0.05 level. *%DIFF* is defined as the effect size, i.e. the "percentage difference of the frequency of a word in the study corpus when compared to that in the reference corpus" (Gabrielatos/Marchi 2012, p. 10). In the analysis, the DiÖ corpus was defined as the study corpus and the Czech corpus as the reference corpus. The negative *%DIFF* indicates underrepresentation in the study corpus.

28 The exact corpus size can be retrieved from the online corpus manager KonText (https://kontext.korpus.cz/, 09. 03. 2018) in the menu that allows users to restrict their search to specific subcorpora.

29 According to Křen et al. (2016, p. 2524) the category of „leisure magazines" is subclassified according to the magazines' main topic: hobby, life style, society, sports or

Tab. 5: Composition of the Czech subcorpus

	tokens per genre	hits per genre
traditional journalism	4 751 614	422
leisure magazines	3 230 133	539

subcorpus, on the other hand, does not contain any source that can be classified as a "leisure magazine", even though most of them contain sections on leisure aspects. Hence, the distribution of the according lemmas in journalistic texts might be related to the subgenre. In order to be able to carry out the test for the second hypothesis, the information given in Tab. 5 is used.

Again, the test proves the hypothesis to hold true: The occurrence of the lemma *zapomínat/zapomenout* 'to forget' is 46.78 % less likely in texts classified as traditional journalism, than in those classified as leisure magazines.[30] A content-related analysis of this tendency remains to be done.

Concerning the present aim, these results imply that due to a bias in the corpus design the results for DiÖ and Czech are not as comparable as originally intended. Still, we argue that this degree of comparability suffices for a first step. If pragmatic aspects of certain constructions of the respective verbs are focused on in follow-up studies, then the sources or corpora need to be comparable with regard to the journalistic subgenre, too.

3.2 Semantic Aspects

While analysing the data, we noticed a remarkable semantic variability of the respective equivalents of 'to forget'. The uses often differ from a presumptive primary meaning, such as 'to be unable to recall or think of something'. Compare the following cases:

(16)   Letos tedy **zapomeňte** *na francouzskou manikúru*.
         'Therefore, forget about the French manicure this year.'
         (SYN2015/Elle, 2/2010)

---

curiosities. Thus, this class of magazines can best be defined as such that primarily deal with a single topic, which is not related to daily news and politics.

30 *%DIFF*: -46.78, *LL:* 95.01; study corpus: "traditional journalism", reference corpus: "leisure magazines".

(17)   […] dass sie <u>eine Pfanne mit Fett</u> auf dem […] E-Herd **vergessen** hatte.
        '[…] that she had forgotten a pan with fat on the stove.'
        (DeReKo-2015-II/Burgenländische Volkszeitung, 13. 01. 2010)

The first case (16) from a Czech fashion magazine, introduces a new fashion norm; the readers are well able to recall or think of the mentioned practice. It is, however, in their best interest, as the author believes, to cease to perform it. The second record (17) from a local Austrian newspaper also does not imply that the agent somehow became mentally incapable of turning off the stove or removing the pan. She merely became distracted and omitted the act required by the circumstances of using a stove.

Such cases are very common in the corpus data, as we show below. One of the hypotheses expects the distribution of the *acc-* and *prep-*construction along semantic lines. Therefore, we attempted to create a suitable annotation system, which allows the capture of different meanings for the equivalents of 'to forget' across languages.

A common approach to the problem – the definition of the meaning of a word – is to consult dictionaries. Jungmann (1839, pp. 525–526) defines two meanings of *zapomenouti* 'to forget': either it means 'to lose one's memory or a capability', or 'to not think about, to not consider something'. Modern German dictionaries like the WDG and *Duden* show a similarly simple definition of *vergessen* 'to forget'. However, whereas *Duden* distinguishes between 'to lose memory of something' and 'to not think about something (anymore)' (cf. Dudenredaktion 2017), the WDG subsumes these aspects under one meaning and contrasts it with the reflexive construction *sich vergessen* 'to forget oneself' (cf. DWDS 2017).

Adelung (1801, col. 1045) distinguishes the meaning of 'to lose a memory' as the "proper" one from the "narrower, figurative" meanings like 'to leave (something) behind', 'to forgive an insult' and 'to make a mistake'. Similarly to Adelung, the DWB (DWB 25, col. 415–424) distinguishes between the proper sense of the verb ('to unconsciously lose a memory', 'to lose a skill due to the lack of exercise'), denoting a "spiritual" (currently, rather "cognitive") activity, from the "playful" and "euphemistic" usages for 'a voluntary refusal to do or commemorate something'. The DWB's definitions thus add a notion of voluntariness as a further dimension to the distinction between 'to lose a memory' and 'to commit a mistake' – or, in the DWB's formulation, 'to not do (something expected)'. This twofold distinction is reflected in the PSJČ (1935–1957), too. This dictionary provides several meanings for the lemma *zapomenout* 'to forget', which can be grouped as follows: a) 'to (involuntarily) lose a memory or skill', b) 'to

(deliberately) disregard a memory', c) 'to (involuntarily) make a mistake' and d) 'to (voluntarily) violate an ethical norm'.

The method of employing categorial definitions of meanings, such as dictionary definitions as annotation categories has been recently employed in studies on cognitive verbs, e.g. by Fabiszak et al. (2014) in their semantic analysis of Pol. *myśleć* 'to think' and its prefixed derivatives. They annotated the attestations with five categories taken from an earlier publication by Danielewiczowa (2002), which offers the following meanings for the verb: "dynamic", "intentional", "knowledge-driven", "factual", or a "hypothetical". Such a system of categories is, however, difficult to use across languages. When we define too few categories, distinctions might get lost, like when we compare examples, such as (17) and (4): Both can be classified as the first meaning given by DWDS 'to lose something from memory', while according to the PSJČ they would reflect different meanings, namely c) (17) or a) (4). On the other hand, when we use too many categories, the question arises as to whether the respective languages share all of them.

For that reason, we prefer not to use a set of particular meanings (or semantic categories), but rather a feature-oriented tagging system, which merely refers to selected aspects of the forgotten object in the phrase. Binary oppositions are used in the mentioned definitions as well, e.g. the cognitive (vs. incognitive or material) character of the object, the voluntariness (or involuntariness) of the action and so on. In the end, two oppositions, one denoting the object's cognitive nature and the other its accessibility were chosen.

With the first feature pair, we mark whether the forgotten object represents a mental content or one of the external world. In cases marked as *cognitive* [+cog], the forgotten objects are memories and skills; the mind is seen as a container of its objects. Here, 'to forget' is synonymous with 'to fail to remember', 'unlearn' or 'disregard'. In other cases, the mind is rather seen as a processor for external stimuli and the forgotten objects exist outside of our mind, like material things (e.g. keys, children) or common values (e.g. fashion, moral conduct). Such concordances are marked as *incognitive* [−cog]. In these cases, the verb 'to forget' is synonymous with 'to neglect' or 'to leave behind'.

The second pair concerns the change in the relation of the forgetting agent to the forgotten object.[31] In the moment of speaking, 'to forget' may denote an

---

31 D'Andrade (1987, p. 115 f.) categorises the semantic classes of cognitive verbs in English along similar lines: on the one hand, the main distinction is made between verbs denoting states and processes, and on the other between verbs denoting achieved states and accomplished actions ('to forget' would be an achieved state-verb).

irreversible change, like being lost in the oblivion or (in case of material things) at some physically remote location. In these cases, in which the change has already happened, the contexts are marked as *inaccessible* [–acc]. The verb is synonymous with 'to not remember' or 'leave behind'. In other cases, which we mark as *accessible* [+acc], the loss implied by the verb 'to forget' is either to be avoided or can be reversed; the object is still accessible to (the mind of) the speaker. In these cases, 'to forget' is synonymous with 'to disregard' or 'to neglect'.[32] Thus, we come to four possible combinations:

[+cog +acc] ('to disregard', 'to not consider')

(18)   a.   A já mu to rád připomínám, když <u>*na* to</u> někdy **zapomene**.
             'And I like to make him remember, when he sometimes forgets about it.'
             (SYN2015/Instinkt 19/2010)

       b.   Man darf nicht **vergessen**, <u>dass jetzt Urlaubszeit war</u>.
             'One should not forget that it was now the holiday season.'
             (DeReKo-2015-II/News, 26.08.2010)

[+cog –acc] ('to not remember', 'to unlearn')

(19)   a.   **Zapomíná** <u>text</u> a svým vzhledem pobuřuje okolí.
             'He forgets the text and incites the surroundings with his looks.'
             (SYN2015/Rytmus života 30/2010)

       b.   **Hans Weigel** […] starb 1991 und wurde sofort <u>vergessen</u>.
             'Hans Weigel […] died in 1991 and was quickly forgotten.'
             (DeReKo-2015-II/News, 23.09.2010)

[–cog +acc] ('to miss to do', 'to neglect')

(20)   a.   **Nezapomene** rozjetá Sigma <u>*na* bránění</u> také v Teplicích?
             'Will not the inflamed Sigma forget about defense in Teplice again?'
             (SYN2015/Mladá fronta Dnes, 27.09.2010)

---

32   The uses of the verb denoting some ethical or intentional problems (e.g. the cases of 'forgetting oneself') are often marked as *accessible*: a value, an obligation, or an intention are still present (i.e. accessible) in the mind, they were merely neglected. A lost memory, a missed opportunity to give a birthday present, or a key left at home is absent, not accessible any more to the speaking subject.

    b.   So mancher **vergaß** beim Zuhören auch <u>die Zeit</u>.
           'In the course of listening, some [spectators] forgot about the
           time.'
           (DeReKo-2017-I/Burgenländische Volkszeitung, 01.07.2010)

[–cog –acc] ('to leave behind'):

(21)   a.   <u>Brýle</u> jsem si **zapomněla**, nic nevidím.
           'I forgot my glasses, I do not see anything.'
           (SYN2015/Respekt 24/2010)

    b.   Vater **vergisst** <u>sein Kind</u> [auf der Tankstelle]
           'Father forgets his child [at the gas station]'
           (DeReKo-2015-II/Burgenländische Volkszeitung, , 26.08.2010)

The concordances were annotated by one of the authors and manually controlled by a second team member. In this process, the semantic features as outlined above proved to be interpersonally valid.

3.3 Annotation Criteria

Besides these semantic features the concordances were manually tagged for selected morpho-syntactic and syntactic features (cf. Tab. 6).

The SYNMOR variable requires further remarks. Both in Czech and in DiÖ, prepositional arguments generally require a correlative element in the matrix clause if realised with a sentence or infinitive phrase. In Czech, these correlating elements are prepositional phrases consisting of a preposition and a demonstrative, e.g. *na to* 'on that'. In the complete Czech sample, only dependent clauses occur after this element (22). Infinitive phrases are not possible.

(22)   Tálibán nebo kdokoli jiný může **zapomenout** <u>*na* to, že odejdeme</u>.
       'The Taliban and everybody else may forget that we will leave.'
       (SYN2015/Právo, 22.11.2010)

Meanwhile, in German (and DiÖ) prepositional arguments require a so-called prepositional adverb, e.g. *darauf* 'on that', as correlates in the matrix sentence. In written texts, such contexts hardly occur. In the complete sample, only a single instance of a dependent infinitive phrase (23) and none of a dependent clause occurs, even though both options are grammatical.

Tab. 6: Annotation criteria

variable abbreviation	description	features
SEMANTIC FEATURES		
SEMCOG	Does the action expressed by the verb refer to a cognitive content?	0 [no]/1 [yes]
SEMACC	Can the object that the verbal action refers to be accessed?	0 [no]/1 [yes]
MORPHO-SYNTACTIC AND SYNTACTIC FEATURES		
SYNCAS	case in which the argument is realised	0 [acc]/1 [prep]
SYNMOR	morphological realisation of the object	NPN [nominal phrase] NPP [pronominal phrase] VPI [infinitive phrase] VPS [dependent clause]
SYNVOI	voice with regard to the whole sentence	*for Czech*: ACT [active] PAS [passive] PSR [reflexive passive construction] *for DiÖ*: ACT [active] PSP [progressive passive] PP* [constructions similar to the progressive passive] PSS [static passive] PS* [constructions similar to the static passive]
SYNFIN	finiteness of the verb in focus	FIN [finite] IFI [plain infinitive] IFM [infinitive as part of a modal verb construction] IFS [infinitive as part of an infinitive construction]
SYNTEM	combination of tense and mood of the verb in focus	INF [infinitive] IFU [indicative future tense] IPS [indicative present tense] IPT [*for Czech*: indicative past tense; *for DiÖ*: indicative preterit] IPF [*DiÖ only*: indicative perfect tense] IPP [*DiÖ only*: indicative plusquamperfect] IMP [imperative] KON [conjunctive]
SYNNEG	sentence negation	0 [pos]/1 [neg]

(23)   […] dass der Leser zwischendurch dar*auf* **vergisst**, ständig zu fragen,
        ob der Attaché nun enthauptet wird […]
        '[…] that in the meanwhile the reader forgets to constantly ask whether
        the attaché will be beheaded […]'
        (DeReKo-II-2015/Falter, 31.03.2010)

Such dependent clauses or infinitive phrases could be analysed as relative clauses of the correlative element. In this context, however, not the surface grammar but the underlying construction grammar is focused upon and thus such examples are annotated as VP-realisations of the *prep*-construction.

4 Contrastive Analysis of Czech and DiÖ: Results and Interpretation

The following contrastive analysis is carried out in three subsequent steps: First of all, the distribution and patterns of variation of the *acc-* and *prep*-construction of Cz. *zapomínat/zapomenout* 'to forget' is modelled with the means of statistical analysis and then described in linguistic terms. Subsequently, we investigate, whether these distributional patterns comply with those of DiÖ *vergessen* 'to forget'. This approach takes the restricted comparability of the analysed data as well as the diverging degree of standardisation of the *prep*-construction in the compared languages into account. Finally, the contexts of the *prep*-construction in DiÖ are analysed with regard to distribution in the various sources and subgenres.

4.1 Distributional Patterns of the Acc- and Prep-Construction in Contemporary Czech

In the first step, the 236 examples in the Czech sample are statistically analysed with a multivariate test in order to predict whether SYNCAS, i.e. the choice of the *acc-* or *prep*-construction is dependent on any of the semantic and syntactic factors. As all the variables are binary or categorical, the test of choice is logistic regression (cf. Field 2009, pp. 264–315; Gries 2012, p. 52).

However, in order to be able to take all annotated features into account in the statistical analysis the annotation system had to be slightly adapted. First of all, the features SYNFIN and SYNTEM had to be combined due to their multicollinearity, i.e. the fact that they were interrelated to a very high degree. Additionally, the SYNMOR and SYNVOI variables were simplified and reduced to simple feature contrasts as shown in Tab. 7.

Tab. 7: Modification of the annotation criteria for the statistical analysis

variable abbreviation	description	features
SYNMOR	morphological realisation of the object	0 [NP realisation]/1 [VP realisation]
SYNVOI	voice with regard to the whole sentence	0 [active]/1 [passive]
SYNFTM	combination of tense and mood of the verb in focus	IFM [infinitive as part of a modal verb construction] IFS [infinitive as part of an infinitive construction] IFU [indicative future tense] IPS [indicative present tense] IPT [indicative past tense] IMP [imperative] KON [conjunctive]

The test was carried out in SPSS v.23 without manipulating the standard settings. The regression model was able to predict 97.9 % of the examples correctly.[33] Tab. 8 shows the results for the independent variables. The single features of SYNFTM are not reported due to their insignificance. The three asterisks (***) indicate high significance of these factors on the $p < 0.001$ level with regard to the dependent variable SYNCAS, i.e. whether the *acc-* or *prep-*construction is chosen.

These results indicate that the independent variables SEMACC and SYNMOR play a significant role in the distribution of the *acc-* and *prep-*construction with Cz. *zapomínat/zapomenout* 'to forget'. The statistical analysis suggests that contexts, in which the direct argument is realised with a sentence or subordinate clause (SYNMOR=1) are less likely to be realised with the *prep-*construction (SYNCAS=1). On the other hand, the use of the *prep-*construction is positively correlated to the accessibility of the object (SYNACC=1).

To discriminate which of these effects is decisive, the logistic regression test was applied in an adapted method (forward step, Wald). This method includes one significant independent variable after the other into the model and thus

---

33 According to Gries (cf. 2012, p. 59) this measure adequately describes the model accuracy if the dependent variable is binary or categorical. The measure usually used to assess a model's accuracy, is Nagelkerke's R^2, which in our case is 0.908 (a measure of 1 would indicate a "perfect model", cf. Field 2009, p. 269).

Tab. 8: Logistic regression of the Czech sample, dependent variable: SYNCAS

	regression coefficient B	standard error	Wald z^2	p
SEMCOG	1.463	1.137	1.566	0.198
SEMACC	8.563	1.750	23.930	0.000 ***
SYNMOR	-8.703	1.571	30.689	0.000 ***
SYNVOI	-3.126	1.637	3.646	0.056
SYNFTM			5.533	0.478
SYNNEG	0.044	1.169	0.001	0.970

allows us to discern to what degree the distribution can be explained by one factor or the other. The model built on the Czech sample was able to predict the dependent variable (SYNCAS) correctly to 89 %[34] by considering exclusively SYNMOR. Adding SEMACC to the model improved the prediction to up to 97.5 %.[35]

Fig. 1 illustrates these findings by showing the absolute distribution of the *prep-* and *acc*-construction for the sample with regard to semantic features and morphological realisation of the direct argument. For this diagram, the original annotations were used (cf. Tab. 6). A first glimpse at the figure reveals the syntactic restriction of the *prep*-construction: Out of 55 examples, in which the direct argument is realised with an infinitive phrase or subordinate clause (SYNMOR=VPS, VPI) all except for three represent the *acc*-construction.[36]

Especially the examples with nominal phrases (NPN) as arguments clearly display the semantic distribution of the *acc-* and *prep*-construction modelled above: Out of the 131 examples, the 118 *prep*-constructions are marked as [+acc] and the 13 *acc*-constructions as [-acc]. The deviations from this pattern among the examples with pronominal phrases (NPP) as arguments may be explained by the fact that these cases require a larger context in order to classify them semantically.

These findings support the preliminary analysis of the complete sample presented by Kim/Scharf (2017): The *prep*-construction is syntactically restricted

---

34 *Nagelkerke's* R^2 = 0.607.

35 *Nagelkerke's* R^2 = 0.902. In a third step, SYNVOI is included (*Nagelkerke's* R^2 = 0.908).

36 In these three cases, the forgotten object is cognitive and accessible, such as in example (20).
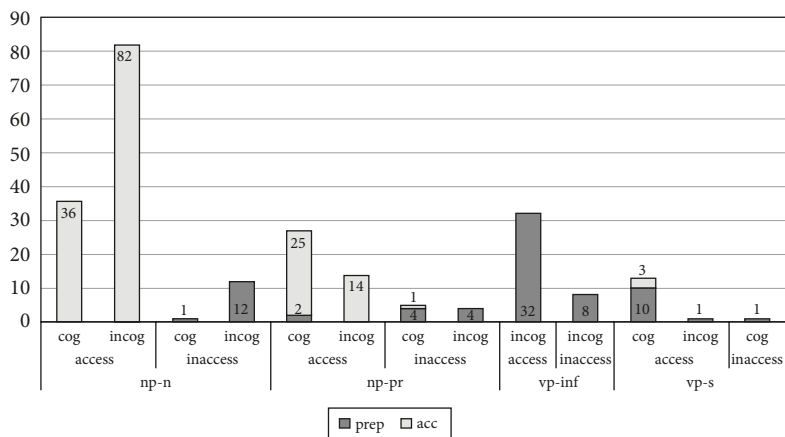
Fig. 1: Distribution of the *acc-* and *prep*-construction in contemporary Czech

to nominal realisations of the direct argument. Semantically, it is clearly linked to those meanings of *zapomínat/zapomenout* 'to forget', in which the object of the act of forgetting is still accessible to (the mind of) the speaker.

4.2 Comparison with the Distribution of the Acc- and Prep-Construction in Contemporary DiÖ

The second step of the distributional analysis verifies, whether the syntactic restriction and semantic distribution that determine the case variation of Czech *zapomínat/zapomenout* 'to forget' apply to DiÖ *vergessen* 'to forget', too. In this context, descriptive statistical methods must suffice, because the degree of comparability necessary for comparative statistical modelling is not given for the two samples because of the following reasons:

(a) First of all, the standardisation degree of the *prep*-construction is different. Therefore, we have to expect the relative frequency of the *prep*-construction to be much lower in DiÖ. This is the case in our sample: Whereas out of 236 analysable Czech examples, 161 (68.22 %) represent the *prep*-construction, only 20 out of 232 examples of the DiÖ sample do so (8.62 %).

(b) Secondly, as outlined above, the corpora are not comparable. We hypothesise that this is reflected in the diverging representation of cognitive and incognitive contexts: In 159 out of the 232 examples (68.53 %) from the
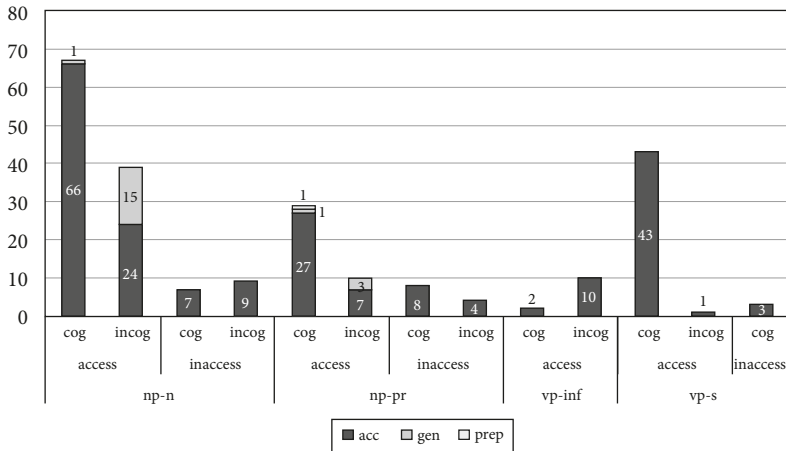
Fig. 2: Distribution of the *acc-* and *prep-*construction in contemporary DiÖ

DiÖ corpus, which exclusively comprises "traditional" journalistic texts, the forgotten object is cognitive. On the other hand, examples with incognitive examples clearly prevail in the Czech corpus; only 83 out of 236 examples (35.17 %) are cognitive. The analysis of the current samples does not allow us to judge, whether in DiÖ the *prep-*construction really occurs more frequently with incognitive objects than in Czech.

Due to these reasons, however, a descriptive analysis must suffice for this second analytical step. Fig. 2 illustrates the distribution of the 232 DiÖ examples along the same semantic and syntactic criteria used for the Czech examples in Fig. 1. The syntactic restriction is evident: not even one out of the 58 examples with VP-argument displays the *prep-*construction. The complete sample, however, contains one (21). Additionally, the semantic distribution modelled for Czech is attested in the DiÖ sample as well: All instances of *vergessen* 'to forget' with a *prep-*construction have an accessible object.

4.3 Detailed Analysis of the Prep-Construction in DiÖ

Due to the semi-standardised character of the *prep-*construction in DiÖ a closer look at the distribution of the *prep-*construction in the different sources is indispensable. Tab. 9 characterises the four sources with regard to their range

Tab. 9: Characteristics of the newspapers and magazines in the DiÖ corpus

source	abbreviation	range[a]	regional distribution	quality
Burgenländische Volkszeitung	BVZ10	– – (3.2 %)	+ + (Burgenland: 44.9 %)	–
Falter	FLT10	– – (1.4 %)	+ (Vienna: 4.5 %)	+
Profil	PRF10	– (6.3 %)	–	+
News	NEW10	+ (10.3 %)	–	–

[a]In comparison to other weekly newspapers and magazines, not including those free of charge. The average range of those publications was 6.5 %.

Tab. 10: Size and (relative) frequency of *vergessen* 'to forget' in the single sources

source	size (words)	total hits (vergessen)	relative frequency of vergessen (pMW)	total hits (prep-constr)	% of the prep-constr. in hits	relative frequency of the prep-constr. (pMW)[a]
BVZ10	4 401 138	237	67.5	40	16.88 %	9.1
FLT10	2 687 057	264	113.5	19	7.19 %	7.1
PRF10	2 501 576	187	81.9	11	5.88 %	4.4
NEW10	1 628 072	174	117.9	7	4.02 %	4.3

[a]Whereas the relative frequency of the lemma can be calculated by and displayed in the corpus query system COSMAS II, the relative frequency of the *prep*-construction was calculated with the tool http://www.thegrammarlab.com/?p=160, 27. 03. 2018.

and regional distribution according to the Austrian Media Analysis 2010[37] and their thematic and linguistic quality.[38] Tab. 10 then shows the distribution of all hits[39] of the lemma *vergessen* 'to forget' in the single sources of the DiÖ corpus and the share of the *prep*-construction amongst them.

In the corpus, the *prep*-construction is more frequent in the sources with a primarily regional range in Eastern Austria, i.e. the *Burgenländische Volkszeitung*

---

37 http://www.media-analyse.at/table/2371, 27. 03. 2018.

38 This assessment mainly represents the authors' personal opinion.

39 This step in the analysis refers to all hits of the lemma *vergessen* 'to forget' in the DiÖ corpus and not only to the annotated ones (cf. table 3).
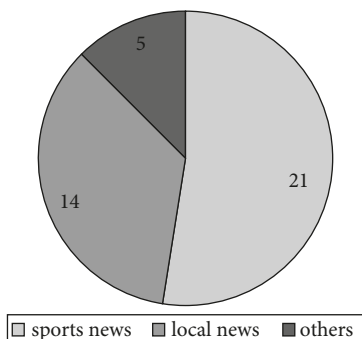
Fig. 3: Distribution of *prep*-constructions in the BVZ-subcorpus

(BVZ) and the *Falter* than in those with an even distribution throughout Austria. Particularly in the BVZ, a regional newspaper specially directed to readers from Burgenland, the *prep*-construction occurs often: Not only is the number of total concordances with the *prep*-construction the largest among all sources (40), but also their share among the total hits of the lemma *vergessen* 'to forget' from the source is almost 10 % higher (16.88 %) than in the source with the second largest share, the *Falter*. Moreover, the *prep*-construction shows the highest relative frequency relative to the whole BVZ-corpus, too (9.1 pMW).

In contrast to the *prep*-constructions from the other three sources, which are distributed across various topics[40] without a certain pattern, the *prep*-constructions from the BVZ display a clear distribution pattern (cf. Fig. 3).

More than half of the 40 examples of the *prep*-construction, namely 21, stem from the sports section. These exclusively represent detailed, narrative descriptions of football matches, i.e. they zoom in on the course of the matches. In twelve cases, the argument is an anthroponym that refers to a (group of) player(s), cf. (24). In the six remaining, it is a deverbal noun, which describes a typical action relevant for the game, such as illustrated by example (2) in the very beginning.

(24)   Die Abwehr **vergaß** bei einem Eckball <u>auf Balasz Czegledi</u> – 1:0.
         'During a corner kick the defence forgot about Balasz Czegledi – 1:0.'
         (DeReKo-2017-I/Burgenländische Volkszeitung, 01.04.2010)

---

40 The topic classification stems from the DeReKo classification and relates to the section
    of the newspaper the article in which it has been published.

Among the 61 *acc*-constructions with *vergessen* 'to forget' from the BVZ sports section, only six occur in similar narrative contexts, cf. example (1). Four of them are morphologically realised with VPs.

These findings suggest that in written, journalistic DiÖ, the *prep*-construction predominantly occurs in regional eastern Austrian sources and in restricted contexts. Unfortunately, the DeReKo does not include any information on the authors of the single newspaper articles, wherefore it cannot be excluded that idiolects of certain journalists play a decisive role. However, follow-up studies on the distribution of the *prep*-construction in DiÖ should focus on both semantic and syntactic, but also regional and pragmatic factors.

5   Conclusion

The findings of our research give evidence that the pattern of case variation of contemporary DiÖ *vergessen* 'to forget' corresponds to the pattern of the equivalent verb in Czech, i.e. *zapomínat/zapomenout* 'to forget'. For both verbs, the direct argument can be realised either with an *acc*-construction or with a *prep*-construction. The respective *prep*-constructions comprise a nominal phrase and functionally very close prepositions, viz. Germ. *auf* 'on', Cz. *na* 'on, at'.

In both languages, these constructions vary along syntactic and semantic lines: First, the *prep*-constructions are largely restricted to noun phrase realisations of the argument, whereas for verbal phrase realisations the *acc*-constructions are preferred. Secondly, the two constructions are linked to specific semantic features of the verb: The *prep*-construction is used in contexts, in which the forgotten object is still accessible to the agent, whereas the *acc*-construction expresses that an object is not accessible anymore.

Moreover, we have shown that the *prep*-construction is indeed restricted to written registers of DiÖ and not common in comparable texts from other parts of the German-speaking areas. However, there is evidence that in Bavaria and some parts of Saxonia, i.e. in regions of potential Slavic-German language contact, the *prep*-construction occurs in spoken sub-standard registers too. Thus, we conclude that the language contact explanation for the phenomenon in focus is plausible.

Sources and References

Dictionaries

Adelung, Johann Christoph. 1793–1801. *Grammatisch-kritisches Wörterbuch der Hochdeutschen Mundart mit beständiger Vergleichung der übrigen*

Mundarten, besonders aber der oberdeutschen. Zweyte, vermehrte und verbesserte Ausgabe. Leipzig. Accessible online: http://woerterbuchnetz.de/cgi-bin/WBNetz/wbgui_py?sigle=Adelung (28. 03. 2018).

Ammon, Ulrich, Rhea, Kyvelos, Rhea, Regula Nyffenegger, & Thomas Oehler. 2004. *Variantenwörterbuch des Deutschen. Die Standardsprache in Österreich, der Schweiz und Deutschland sowie in Liechtenstein, Luxemburg, Ostbelgien und Südtirol*. Berlin, New York.

Ammon, Ulrich, Hans Bickel & Alexandra N. Lenz (eds.). 2016. *Variantenwörterbuch des Deutschen. Die Standardsprache in Österreich, der Schweiz, Deutschland, Liechtenstein, Luxemburg, Ostbelgien und Südtirol sowie Rumänien, Namibia und Mennonitensiedlungen. 2., völlig neu bearbeitete und erweiterte Auflage*. Berlin.

Bergmann, Gunter & Dagmar Helm (eds.). 1994–2003. *Wörterbuch der obersächsischen Mundarten*. Berlin.

BWB = Bayerische Akademie der Wissenschaften (eds.). 1995‑2016. *Bayerisches Wörterbuch*. München [not completed].

DWB = Grimm, Jacob & Wilhelm Grimm (eds.). 1854‑1961. *Deutsches Wörterbuch von Jacob und Wilhelm Grimm. 16 Bde. in 32 Teilbänden*. Leipzig. Accessible online: (28. 03. 2018).

DWDS = Digitales Wörterbuch der deutschen Sprache. 2017. „vergessen", in: *Wörterbuch der deutschen Gegenwartssprache (1964–1977)*, kuratiert und bereitgestellt durch das Digitale Wörterbuch der deutschen Sprache. Accessible online: https://www.dwds.de/wb/wdg/vergessen (10.12.2017).

Dudenredaktion. 2017. „vergessen" auf Duden online. Accessible online: https://www.duden.de/node/706902/revisions/1656620/view (10.12.2017).

Engels, Heinz, Norbert Englisch & Ernst Schwarz (eds.). 1982–2010. *Sudetendeutsches Wörterbuch: Wörterbuch der deutschen Mundarten in Böhmen und Mähren-Schlesien*. München.

Jungmann, Josef. 1839. *Slovník česko-německý, Díl V., V–Z*. První vydání. Praha.

Maurer, Friedrich & Rudolf Mulch (eds.). 1965–2010. *Südhessisches Wörterbuch*. Marburg.

Ochs, Ernst & Karl Friedrich Müller (eds.). 1925–2009. *Badisches Wörterbuch*. Lahr, Wien.

ÖWB = Pabst, Christiane M., Herbert Fussy & Ulrike Steiner (eds.). 2016. *Österreichisches Wörterbuch. 43. Auflage*. Wien.

PSJČ = Hujer, Oldřich, Emil Smetánka, Miloš Weingart, Bohuslav Havránek, Vladimír Šmilauer & Alois Získal (eds.). 1935–1957. *Příruční slovník jazyka českého*. Praha. Accessible online: http://psjc.ujc.cas.cz/ (31.03.2018).

Schaechter-Viswanath, Gitl & Paul Glasser. 2016. *Comprehensive English-Yiddish Dictionary (based on the lexical research of Mordkhe Schaechter)*. Bloomington, Indianapolis.

SSJČ = Havránek, Bohuslav & Kol. 1960–1971. Slovník spisovného jazyka českého. Praha. Accessible online: http://ssjc.ujc.cas.cz/ (31.03.2018).

WBÖ = Österreichische Akademie der Wissenschaften (eds.). 1970‒2015. *Wörterbuch der bairischen Mundarten in Österreich*. Wien [not completed].

Zehetner, Ludwig. 2014. *Bairisches Deutsch. Lexikon der deutschen Sprache in Altbayern. 4., überarb. und erw. Aufl*. Regensburg.

Corpora

Benko, Vladimír. 2014. Aranea: Yet Another Family of (Comparable) Web Corpora. In: Petr Sojka, Aleš Horák, Ivan Kopeček & Karel Pala (eds.): *Text, Speech and Dialogue. 17th International Conference, TSD 2014, Brno, Czech Republic, September 8-12, 2014. Proceedings. LNCS 8655*. Basel.

CMY = The Corpus of Modern Yiddish. Accessible online: http://web-corpora.net/YNC/search/ (09.04.2018).

Institut für Deutsche Sprache. 2015. *Deutsches Referenzkorpus/Archiv der Korpora geschriebener Gegenwartssprache 2015-II* (Release vom 28.09.2015). Mannheim: Accessible online: www.ids-mannheim.de/DeReKo (28.03.2018).

Křen, Michal, Václav Cvrček, Tomáš Čapka, Anna Čermáková, Milena Hnátková, Lucie Chlumská, Tomáš Jelínek, Dominika Kováříková, Vladimír Petkevič, Pavel Procházka, Hana Skoumalová, Michal Škrabal, Petr Truneček, Pavel Vondřička, & Adrian Jan Zasina (eds.). 2015. *SYN2015: reprezentativní korpus psané češtiny*. Praha 2015. Accessible online: http://www.korpus.cz (28.03.2018).

Rosen, Alexandr, Martin Vavřín & Adrian Jan Zasina. 2017. *The InterCorp Corpus: Belorussian, Bulgarian, Czech, Macedonian, Polish, Russian, Serbian, Slovak, Slovenian, Ukrainian. Version 10 of 09. September 2017*. Institute of the Czech National Corpus, Charles University, Prague, Accessible online: http://www.korpus.cz (09.04.2018).

Literature

Blahak, Boris. 2015. *Franz Kafkas Literatursprache. Deutsch im Kontext des Prager Multilingualismus*. Köln.

Břenek, Oldřich. 2017. *Sprachgebrauch und Sprachbeurteilung in Österreich am Beispiel der jüngeren Generation*. Berlin.

Cvrček, Václav & Olga Richterová (eds.). 2018. en:cnk:syn2015, In: *Příručka ČNK*. Accessible online: https://wiki.korpus.cz/doku.php?id=en:cnk:syn2015&rev=1481470218 (09.03.2018).

D'Andrade, Roy. 1987. A folk model of the mind. In: Dorothy Holland, & Naomi Quinn (eds.), *Cultural models in language and thought*. Cambridge, 112–148.

Danielewiczowa, Magdalena. 2002. *Wiedza i niewiedza: Studium polskich czasowników epistemicznych*. [Knowing and not knowing: A study of Polish epistemic predicates.] Warszawa.

Eichhoff, Jürgen. 1993. *Wortatlas der deutschen Umgangssprache*. Vol. III. München.

Fabiszak, Małgorzata, Anna Hebda, Iwona Kokorniak & Karolina Krawczak. 2014. The semasiological structure of Polish *myśleć* 'to think'. A study in verb-prefix semantics. In: Dylan Glynn & Justyna A. Robinson (eds.), *Corpus Methods for Semantics. Quantitative Studies in Polysemy and Synonymy*. Amsterdam, Philadelphia, 223–251.

Field, Andy. 2009. *Discovering statistics using SPSS. Third edition*. Los Angeles et. al.

Friebertshäuser, Hans. 1983. Die großlandschaftlichen Wörterbücher der deutschen Dialekte. Areale und lexikologische Beschreibung. In: Werner Besch (ed.), *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung. 2. Teilband (= Handbücher zur Sprach- und Kommunikationswissenschaft, 1.2.*) Berlin, New York, 1283–1295.

Gabrielatos, Costas & Anna Marchi. 2012. Keyness: Appropriate metrics and practical issues. Presentation held at: *CADS International Conference 2012. Corpus-assisted Discourse Studies: More than the sum of Discourse Analysis and computing?, 13–14 September 2012, University of Bologna, Italy*. Accessible online: http://repository.edgehill.ac.uk/4196/1/Gabrielatos%26Marchi-Keyness-CADS2012.pdf (09.04.2018).

Gries, Stefan Th. 2012. Statistische Modellierung. In: *Zeitschrift für germanistische Linguistik 40*, 38–67.

Jedlička, Alois (1990): Tschechische Lexikographie. In: Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand & Ladislav Zgusta (eds.), *Wörterbücher. Ein internationales Handbuch zur Lexikographie. 2. Teilband (= Handbücher zur Sprach- und Kommunikationswissenschaft, 5.2)*. Berlin, New York, 2278–2284.

Kretschmer, Paul. 1918. *Wortgeographie der hochdeutschen Umgangssprache*. Göttingen.

Kim, Agnes. 2019. Multilingual Lower Austria. Historical sociolinguistic investigations on the Wenker questionnaires. In: Lars Bülow, Ann-Kathrin Fischer & Kristina Herbert (eds.), *Dimensionen des sprachlichen Raums: Variation – Mehrsprachigkeit – Konzeptualisierung*. Wien, 187‑211.

Kim, Agnes & Sebastian Scharf. 2017. Areal Variation in Case Government of the German verb vergessen 'to forget' in contrast with Czech zapomínat/zapomenout 'to forget'. Poster presented at Linguistics Prague 2017, 27.04.2017‑29.04.2017, Charles University. Accessible at: https://dioe.at/details/artikel/565/ (28.03.2018).

Kühn, Peter & Ulrich Püschel (1990): Die deutsche Lexikographie vom 17. Jahrhundert bis zu den Brüdern ausschließlich. In: Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand & Ladislav Zgusta (eds.), *Wörterbücher. Ein internationales Handbuch zur Lexikographie. 2. Teilband (= Handbücher zur Sprach- und Kommunikationswissenschaft, 5.2)*. Berlin, New York, 2049‑2077.

Křen, Michal, Václav Cvrček, Tomáš Čapka, Anna Čermáková, Milena Hnátková, Lucie Chlumská, Tomáš Jelínek, Dominika Kováříková, Vladimír Petkevič, Pavel Procházka, Hana Skoumalová, Michal Škrabal, Petr Truneček, Pavel Vondřička, & Adrian Jan Zasina. 2016. SYN2015: Representative Corpus of Contemporary Written Czech. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portorož, 2522–2528.

Muhr, Rudolf. 1995. Grammatische und pragmatische Merkmale des österreichischen Deutsch. In: Rudolf Muhr, Richard Schrodt, Richard & Peter Wiesinger (eds.), *Österreichisches Deutsch. Linguistische, sozialpsychologische und sprachpolitische Aspekte einer nationalen Variante des Deutschen*. Wien, 208–234.

Newerkla, Stefan Michael. 2007. Areály jazykového kontaktu ve střední Evropě a německo-český mikroareál ve východním Rakousku. In: *Slovo a Slovesnost* 68, 271–286.

Rayson, Paul & Roger Garside. 2000. Comparing corpora using frequency profiling. In: *Proceedings of the workshop on Comparing Corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000). 1‑8 October 2000*. Hong Kong, 1‑6.

Scharf, Sebastian. 2018. *Variation in der Kasusrektion der Äquivalente für ‚to forget' in zentralost- und osteuropäischen Sprachen*. Berlin. MA thesis.

Schuchardt, Hugo. 1884. *Slawo-Deutsches und Slawo-Italienisches. Dem Herrn Franz von Miklosich zum 20.11.1883*. Graz.

Teweles, Heinrich. 1884. *Der Kampf um die Sprache. Linguistische Plaudereien*. Leipzig.

Wiesinger, Peter. 2015. Das österreichische Deutsch in der globalisierten Umwelt: Wandlungen durch bundesdeutsche Einflüsse. In: Alexandra N. Lenz (ed.), *Dimensionen des Deutschen in Österreich: Variation und Varietäten im sozialen Kontext*. Frankfurt/Main, 91–122.

Zeman, Dalibor. 2003. *Das österreichische Deutsch und die österreichisch-tschechischen Sprachbeziehungen. Ein kulturhistorischer und sprachlicher Abriß*. Dissertation. Universität Wien.

František Martínek

Remarks on the Development of the Czech Modality System in Contact with German[1]

Abstract: The present paper offers five diachronic case studies which are devoted to the formation and development of several Czech modal expressions, taking into account the German influence on the Czech modality system. In the studies about the verbs *mít* 'should' and *potřebovat* 'to need', which are used as a conditional marker and as a deontic modal verb respectively, two semantic loans of peripheral modal meaning are demonstrated. The studies about *hodlat* 'to want (to do sth.)' and *uspět* 'to succeed in sth.' as well as the modal adverb *možná* 'maybe', exemplify the diachronic development to and from a modal meaning/function. In the last study, semantic parallels between the Czech modal particle *také* and German *auch* 'also, too, as well' are explained by the diachronic relationship of both languages.

Keywords: language contact, Czech, German, modality, diachrony

1 Introduction

This contribution closely focuses on several particular topics relating to both Czech and German modality. It does not deliver a detailed overview on the whole subject, but introduces five separate studies as examples for the development of modality means. These individual issues are connected to some general information about the Czech modality system and also to previous research results.

The paper is structured as follows: The first three sections offer a theoretical background on modality. Section 1, the introduction, outlines some characteristics of the diachronic and areal approach and provides basic information about German-Czech language contact. Section 2 refers to seminal research on modality in Czech and mentions some arguments for the influence of German on the Czech modality system. The following five sections 3-7 are case studies of a number of interesting diachronic topics: *mít* 'should' as a conditional marker (section 3), *potřebovat* 'to need' as a deontic modal verb (section 4), the opposite development to and from a (quasi) modal verb demonstrated by the examples *hodlat* 'to want (to do sth.)' and *uspět* 'to succeed in sth.' (section 5), the

formation of the modal adverb *možná* 'maybe' (section 6), and the diachronic relationships between Czech *také* and German *auch* 'also, too, as well' (section 7). The final section 8 summarizes the partial results and provides an outlook for contemporary Czech.

1.1 Terms and Definitions

In the beginning, an appropriate and—if possible—comprehensible terminology will be elaborated. This paper focuses on two types of expressions—first on modal verbs and second on some groups of non-inflectional words, which are usually referred to as modal adverbs, sentence particles, modal particles and the like.[2] Marginally, one will find references to other constructions expressing modality, e.g., *být k dostání* 'to be available'.

Among the verbs relating to modality, one can distinguish—and certainly not only in German and Czech—a small group of *canonical modal verbs*, e.g., Czech *muset* 'must, to have to', from another group. Let us call the latter group's members *modal verbs in a wider sense* or *quasi-modal verbs*, e.g., Czech *hodlat* 'to want (to do sth.)', which exists besides the synonymous canonical modal verb *chtít*. From a diachronic perspective, one can register movements between these two groups, as will be demonstrated in section 5. Furthermore, several *modal analytical predicates* such as Czech *mít zájem* 'to have interest, to be interested' are to be counted among the verbs relating to modality.

All non-inflectional words expressing modality will hereafter be referred to as *modal words*. In sections 7 and 8, two different types of modal words will be analyzed. The first of them, *možná* 'maybe', expresses the degree of the speaker's certainty. It will be called *modal adverb* in this function further on. The second word, *také* 'also, too', is a prime example for polyfunctionality, and this paper focuses on one of its functions only, namely, the function of a *modal particle*.

1.2 Corpora and Other Resources

It is clear that new electronic sources containing historical Czech texts make it possible to describe small shifts in the development of Czech and, thus, to understand language change. However, this optimistic statement has to be refined with regard to two aspects: First, besides the corpora of older texts (*Staročeská textová banka*, *Diakorp*, *KH-dopisy*, *19th-Century Archive*), many electronic dictionaries,

---

2   This paper does not pursue a broad theoretical discussion about the significantly difficult delimitation of adverbs and particles, which is based on the semantics and functions of the related expressions in various contexts.

digitized word indexes and card indexes (so called "lexical archives", e.g., *Lexikální databáze humanistické a barokní čeština* and the *Lexical Archive of the PSJČ*, among others) can also be considered new electronic sources.[3] The work with examples hand-written on cards and which, in many cases, have to be checked against the original sources, does not fully correspond to the idea of modern linguistic work, but due to the *small size of Czech historical corpora*, this work is still indispensable up to now. Second, this paper presents some quite peripheral, i.e., *less frequent phenomena*, for which a very small number of occurrences is available in the corpora. Hence, this could also serve as motivation for broadening and extending the existing corpora. The analysed examples will demonstrate the possibilities but, at the same time, the paucity and limitations of the existing resources—corpora and dictionaries. Finally, this paper will also suggest sources one can use if corpora have too little data for specific research purposes or, in the worst case, an appropriate corpus does not exist at all.

1.3 Diachronic and Areal Approach

For all the topics under discussion, the diachronic and the areal approaches will be used. The *diachronic approach* to the data enables us to problematize simple yes/no statements. This is necessary, since the data stems from very different time periods, going back in the course of history: from contemporary Czech to the language of the 19th and 20th century. In the course of these 200 years, Czech language history was significantly shaped by two factors, namely, the Czech

---

3   *Staročeská textová banka* is a corpus of Czech texts prior to 1500, currently consisting of about 5.3 million tokens. *Diakorp* is a corpus of Czech texts from the oldest times up to the 20th century, comprising about 4 million tokens in the current version (v. 6). The *19th-Century Archive* is an extensive, but unfortunately non-public corpus of Czech 19th-century texts available at the Institute of the Czech National Corpus in Prague. *KH-dopisy* is a small corpus of all preserved letters written and received by the Czech writer and journalist Karel Havlíček (1821–1856), totalling 622 thousand tokens. *Lexikální databáze humanistické a barokní čeština* consists of up to 550 thousand excerpts of words from Middle Czech (approximately 1500 to 1780). The *Lexical Archive of the Dictionary PSJČ* (of 8.7 million excerpts) consists of Czech words from the Czech National Revival to the 1990s, i.e., it was being further supplemented after the completion of the PSJČ (*Příruční slovník jazyka českého* = Concise dictionary of the Czech language) in the 1950s and used for the two following Czech dictionaries, SSJČ (*Slovník spisovného jazyka českého* = Dictionary of the Standard Czech language) and SSČ (*Slovník spisovné čeština pro školu a veřejnost* = Dictionary of Standard Czech for schools and the public). Details about all these data sources are given in the literature at the end of this article.

National Revival in the 19th century and the expulsion of the Czech Germans after World War II. In some cases, even data from the Old Czech, i.e., the medieval Czech, and the Middle Czech, i.e., the Czech of the Early modern period, is considered. Hence, it is hard to claim that a certain phenomenon does/does not exist in Czech. Not only the non-standard language varieties, but also the history of the language must be taken into account.

With regard to the modality system, the diachronic approach is compatible with the following concept: Languages contain lexical units, i.e., both individual words and phrases, with a certain potential to turn into words able to express modality.[4] The crucial research question then is: What factors cause these shifts and changes in the meaning and function of lexical units?

From the *areal approach* to Czech-German language contact follows, that geographically close, i.e., neighbouring Czech and German dialects from the same time period need to be examined. This is especially necessary since German was—and still is—a far more diversified language than Czech regarding its horizontal variation. Special attention has to be paid to parallels of Czech and Bavarian dialects of German, and Austrian German in particular.[5]

A related question is the direction of influence, in which the lexis differs from the grammar. Several Czech loanwords and a few loaned phrasemes can be found in the German dialects used in regions neighbouring the Czech lands. Conversely, German has evidently influenced Czech in the majority of its grammatical innovations. The German origin of some Czech innovations in the field of modality will be exemplified below.[6]

1.4 Language Contact

In the following studies, two groups of factors will be considered: internal factors, i.e., the development of a language on its own, and external factors, i.e., developments caused by language contact. Other factors, e.g., sociolinguistic ones, can be omitted here. A classic definition of the language contact effect reads as follows:

---

4   Phrases with modal function are called *phrasal items* or *modal idioms*, amongst others.
5   Cf. Newerkla (2007) on Czech influence on German in Vienna and Eastern Austria.
6   It is beyond the scope of this paper to draw parallels in Central European languages other than Czech and German as well as to discuss the terminology of the *Central European area*, *mitteleuropäischer*/*Central European Sprachbund*, *Donausprachbund* etc.; more on this topic cf. Januška, this volume.

[A]n external explanation for a particular structural change is appropriate, either alone or in conjunction with an internal motivation, when a source language and a source structure in that language can be identified. The identification of a source language requires the establishment of present or past contact of sufficient intensity between the proposed source language and the recipient language. (Thomason/Kaufman 1988, p. 63; quoted as in Berger s. a., p. 17)

To this generally accepted definition, we may add the following: One has to consider that the *internal and external motivations* do not necessarily stand in direct opposition. On the contrary, both can be closely connected and therefore support each other. For example, a specific language change can be triggered by a foreign impulse and then carry on to the target language on its own. Moreover, one has to distinguish between the results of language contact of two or more languages that are geographically close (neighbouring), and the following phenomena:

a) correspondences caused by the *genetic relationship* of the languages,
b) the so-called *Europeisms*, and
c) *system-based, 'regular' shifts in meaning and/or function* of lexical items, without the necessary external motivation.

The *Europeisms* appear in a large number of European languages, mostly in those which were strongly influenced by Latin—and partially by Greek—in some period(s) of their development. On the other hand, the role of German in the transfer of these—primarily lexical—phenomena into Czech cannot be denied.

The system-based shifts can be caused by an analogy or following a frequently repeated process, among others, and without influence from other languages. Two examples can be offered here, the second one from the field of modality. The metaphors like *right (hand)* 'good, strong' vs. *left hand* 'bad, weak' are common in most European languages. Many languages also display the following development scheme: an item which already expresses deontic modality tends to additionally express epistemic modality.

In his comparative studies of German and Polish modal verbs and modal verbs, Weiss (1987; 2009) points at some more issues that require consideration with regard to language comparison and language contact. His findings can be paraphrased in the following four points—which may sound trivial, but are certainly not self-evident:

a) If an expression (a word, phrase, or multi-word item) has parallel meanings in the two languages that are being compared, it is reasonable to examine it in relation to language contact. However, there is still no guarantee that the parallel meanings have actually resulted from language contact.

b) Later evidence of a certain parallel meaning in a language may constitute strong proof of the language contact effect and also of the direction of the influence.

c) Another type of solid evidence of the direction of the influence can be the number of word meanings. It tends to be higher in the source than in the target language, because although the meaning structure in the target language has been formed under the influence of the source language, it usually does not adopt all the meanings the word possesses in the source language.

d) The grammaticalization processes, which are repeated in many (not closely related) languages, are to be evaluated as a potentially strong argument against the language contact effect (cf. Weiss 2009, p. 131).

2 Previous Research in the Field of Modality and German-Czech Language Contact

2.1 German Influence on Grammatical Phenomena on Czech in General

Before discussing the topics of language contact in detail, and connecting them to some examples of the empirical research on modality in Czech, a more theoretical approach should be mentioned. In his paper "The Influence of German on the Grammatical System of Czech" (Berger s. a.; German version: Berger 2008), Berger discusses all the grammatical phenomena ever mentioned in any work from Czech linguistics—including the "partially grammaticalized"—as possibly having a German origin or having been influenced by German. He sorts these phenomena into five groups, which, in addition to evident contact phenomena and phenomena which clearly cannot be traced to language contact, include three forms of the possible support and influence of German in Czech language development:

1) clear contact ("phenomena which can be traced back to language contact with high probability"),

2) "phenomena for which language contact may have been the motivating factor, but which then led to a similar, yet independent development",

3) "phenomena which should be seen within the context of a larger area (i.e. within a Western European context), although they may have entered Czech through German"

4) "phenomena which should be explained as general developments in modern standard languages", and

5) the exclusion of any German influence (Berger s. a., p. 18).

In this context, it is important that Berger categorizes all phenomena concerning modality into the first group. This regards the modal verbs, the modal construction *být k dostání* 'to be available' (lit. 'to be to get'), and "probably also the 'new' [Czech] particles" (Berger s. a., p. 18), beside other modal particles. In his opinion, they are in all probability the result—or product—of German influence.

2.2 Seminal Research on Modality in Czech

Let us have a brief look at several concrete examples of previous inquiries regarding Czech modal verbs and modal words (as defined in 1.1). There exist detailed descriptions of various Czech modal constructions of a finite verb and an infinitive, expressing *volitive modality*. According to a rather old delimitation by the Czech linguist Grepl from the 1970s, this category includes *intention*, *necessity*, *possibility* and *ability* (cf., e.g., Grepl 1973 and PMČ 1995, pp. 533–547 and 627–630). For example, Karlík and Štícha describe the infinitive in a syntactic structure with verbs *být* 'to be' and *mít* 'to have' expressing possibility (*Mám/Je kde spát* 'I have/There is a place to sleep'; Karlík/Štícha 2011: pp. 941–944). Kolářová (1999) outlines the shift in meaning of several verbal forms of *(po) dívat se*, *hledět* and *koukat (se)*, all meaning 'to look (at sth.)'. Some of them are used as a modal verb and obtained the meaning 'let's do …'; and some, e.g., the imperatives *hleď*, *dívej* '(let's have a) look', have developed into modal particles expressing an appeal.

In Hansen/Nekula/Banášová (2011), the authors elaborate on the construction present in the sentence *Karla Gotta nemusím* 'I do not like Karel Gott (very much)' and outline a new meaning of the verb *muset* (lit. 'must, to have to'), namely 'to not like'. This meaning is connected to a special syntactic construction with negation and its development can be explained as a possibly incipient decline of a modal verb to a full verb.

Finally, an example for research on modal particles will be given: Fried (2007) outlines the shift of the conditional conjunction *jestli* 'whether' to a particle expressing doubts. This particle can be complemented with *jen* (lit.: 'only'), forming a multi-verb particle expression, or emphasized with words like *vůbec* 'at all, (not) ever', cf. the following examples of direct speech which are analogous to examples by Fried (2007, pp. 62 and 63): *Jen jestli tu práci udělal!/Jestli tu práci vůbec udělal!* both meaning 'I doubt he ever finished the work'.

2.3 Arguments for the Influence of German on the Czech Modality System

Several convincing arguments for the influence of German on the Czech modality system have been proposed:

1) Among Slavic languages, Czech has the greatest number of modal verbs in its repertoire. The following are, among others, two differences between Czech and the geographically close Polish: The first one is the function of a complete verbal paradigm of *mít* (lit.: 'to have') in Czech as a counterpart of the English *should* in comparison to the frozen verbal form *powinien* in Polish (cf., e.g., Weiss 1987). A second difference is the Czech development of the verb *smět(i)* 'to be allowed to': its modal meaning developed from the Old and Middle Czech full verb *smĕti* 'to dare (to do sth.)' which retains the original meaning in Polish as *śmieć*.

2) A very solid argument for German influence is the use of deontic modal verbs for the expression of epistemic modality, like *Má tam jít* 'He should go there (because he was forced to do it)' → 'I think that he should go there.' Of course, the mere existence of parallel meanings in German and Czech cannot be considered sufficient proof of German influence. However, it is attested that in Czech this development took place later than in German (for this argumentation, based on parallel Polish examples, cf. Weiss 2009).

3) Czech purists traced the construction of the verb *být* 'to be' with the preposition *k* 'to' and a verbal substantive ending *-ní/-tí*, like *být k dostání* 'to be available', lit.: 'to be to get', to German influence, since it corresponds to German *zu bekommen sein* (cf. Weiss 1987 and Berger s. a., pp. 7 and 18, where this is "with high probability" classified as a product of language contact). This modal construction can express possibility, as shown, as well as ability or necessity in other cases. Of particular interest is the position of the Czech (rather anti-purist) linguist V. Ertl, who rejected the German influence because of older Czech evidence of this construction, as well as a lack of total correspondence between the Czech and the German form. According to him, a Czech speaker should reportedly have recognized *zu bekommen* as an infinitive and would therefore not have translated it with a verbal noun (Ertl 1927, pp. 226–227). In any case, this speculation does not seem conclusive, since it overestimates the analytical capabilities of language users.

4) The last example of German influence on Czech is the domain of the Czech modal particles (German *Abtönungspartikeln*, among other terms). Their development in Czech shows a considerable amount of correlation with their German counterparts:

In spite of some differences in the norms of the individual Czech and German particles, the results of this inquiry demonstrate the far reaching and interesting threefold equivalence between the particle systems of Czech and German. This equivalence is of even greater interest because Czech differentiates itself in this from the other Slavic languages. (Nekula 1996, p. 196)

Moreover, M. Nekula argues for the restructuring of the original "Slavic" Czech particle system under the influence of the German system.

The following empirical part of the paper also documents many of the tendencies described above. Let us begin with two specific modal functions of the verbs *mít* 'should' and *potřebovat* 'to need'.

3 *mít* 'should' as a Conditional Marker

In his study of German-Polish relationships in the modality domain, Weiss surveys the correspondence between the conjunctive II of German *sollen* (*ich sollte* 'I should', *du solltest* 'you should', …; further referred to by the form of the 1st and 3rd person singular *sollte*) and the Polish verb *mieć* 'should' which both are used to mark a condition (Weiss 2009, p. 138). Weiss argues that Polish does not follow the German pattern of a dependent conditional clause without conjunction, in which the form *sollte* (and not the conditional verbal form with *by* like in Polish) expresses the condition. In Polish, only a dependent clause with a conjunction and a conditional verbal form is acceptable:

(1)   a.   *Sollte* ich mich *geirrt haben*, so nimm es mir bitte nicht übel.
      b.   *Miałbym się mylić, nie miej mi tego za złe.
      c.   Gdybym się mylił, nie miej mi tego za złe.
           'If I have been wrong [lit.: should I have been wrong], please do not take it amiss.'
           (Weiss 2009, p. 138, example nr. 15; original emphasis, English translation F.M.)

The situation in Czech evidently differs from the one in Polish. Sentences with *mít* 'should' expressing a condition seem to be acceptable to contemporary speakers, albeit with a conjunction and a conditional verbal form only (e.g., *když … by*, contracted into *kdyby*). A dictionary from the 1960s contains the example *kdyby se ti to nemělo líbit, vrať to* 'if you do not like it, give it back' (SSJČ, s.v. *míti* 10c).

The following observations do not concern contemporary Czech. They only support the claim that the Czech modality system is closer to German than the modality

system in other Slavic languages (cf. point 2.3). The relatively frequent occurrence in informal texts dating to the middle of 19th century, e.g., in the correspondence of Karel Havlíček (*KH-dopisy*), suggest that we encounter a contact phenomenon.

In the examples of clear conditional use, a substitution with *muset* 'must, to have to' would change the meaning of the sentence:

(2)   a.   *Kdyby jste* ale *neměl* brzo k nám *přijedst*, pište, jaké knihy mám nakúpit […].
             'If you do not come to us soon, write to me what books I should buy […].
             (1845-02-28, Josef Henzl; emphasis here and in all following examples F.M.)

       b.   […] proto *kdybychom se rozejít měli* za příčinau zmařené föderaci aby někdo z Magnatů českých s námi byl.
             '[…] because if we break up due to the failure of the federation plans, may there be some Bohemian tycoon on our side.'
             (1849-01-27, František Brauner)

       c.   *Kdyby* tedy třeba jen tato okolnost *vaditi měla* a nebylo jiných překážek […].
             'If only this circumstance poses a hindrance and there are no other obstacles […].'
             (1851-07-17, Karel Havlíček)

Sometimes the conditional use appears as a concession, i.e., a negative condition. In some of these examples, it is particularly difficult to differ between the conditional use of *mít* and the expression of the duty. Although the ambiguous examples (3a) and (3c) can be read in this sense of a duty, in which case the verb *mít* can be replaced with *muset* 'must, to have to' and the meaning can be interpreted as a writer's future plan, I prefer the conditional interpretation 'if a circumstance would occur'. In any case, these transitional examples show us the developmental tendencies and possible innovations of the modal domain.

(3)   a.   A v restu zůstat nesmím a nechci, *kdybych měl prodat* všechno, co mám […].
             'And I must not and do not want to stay in debt even if I were to sell everything I own.'
             ([1847-11-10], Karel Havlíček)

b. Při tom se ale, *kdyby se* to i *podařiti mělo*, obávám řečí lidských, neboť
se jistě najdou hloupí pleskalové […].
'Even if this succeeds, I still fear gossip, because dumb babblers will
surely appear […].'
(1851-07-19, Karel Havlíček)

c. […] *i kdybych měl* v nejhorším pádu si na to [= knihy] v Brodě
pokojík *najmout*
'[…] even if I were, in the worst case, to hire a room for them [books]
in Brodʼ
(1853-03-01, Karel Havlíček)

The following examples in (4) also seem to be ambiguous, falling in between
the conditional use and the expression of duty. The replacement of *mít* by *muset*
comes into consideration, but it is connected with a disambiguation of the
meaning.

(4)   a.   Žes milý Karle ošemetný lhář, to může i slepý vidět; *kdyby byl měl*
p. Vikář a Sekretář na přivitáni a poděkováni Redaktora […] *čekat*, ti
by se načekali […].
'You are, dear Karel, a tricky liar, even a blind man can see that; if the
Messrs. Vicar and Secretary were to wait for the welcome and thanks
of the editor of the journal, for you, they would have to wait forever
[…].'
(s. a., František Mudra)

b.   Nevěříte, že bych si musel 3 páry ukrajinských volů připřáhnout k
ruce, *kdybych* to *měl poslat*!
'You do not believe that I would have to harness three pairs of
Ukrainian oxen to my hand, if I were to send it!'
(1844-06-06/05-22, Karel Havlíček)

c.   *Kdybych* ti *měl psát*, co dělam, co pracuju, nepopsal bych tím mnoho
papíru […].
'If I had to write you about what I'm doing, what I'm working on,
I would not fill many sheets of paper in doing so.'
(1845-10-02, Gabler Vilém)

From the abovementioned examples, two questions emerge for future research:

a) This modal use nowadays seems peripheral in Czech—this impression has to be verified by future corpus research—and it appears to have been more frequent in the 19th century. One can suppose that this phenomenon is weakening due to little or no contact with German. However, it has to be kept in mind that this construction may belong to informal spoken language, where there are no sufficient records from the older times.

b) The assumption, that sentences beginning with *Měl bych…* are unacceptable for a certain part of contemporary Czech speakers can also be verified by corpus or questionnaire research.

A very similar development as with the conditional *mít* 'should', namely a supposed regression/elimination of another modal construction which does not have support in German-Czech language contact anymore, will be demonstrated in the next section.

4 *Potřebovat* 'to need' as a Deontic Modal Verb

Weiss claims (2009, p. 134) that only the *deontic* use is attested for the Polish verb *nie potrzebować* and the Czech *nepotřebovat*, a counterpart to German *nicht brauchen* (all meaning 'to not need'). He claims that no evidence of its epistemic use occurs, which, however, is possible with the German equivalent *nicht brauchen*.

This statement corresponds to a broad description by Hansen (2001). This author also excludes the epistemic use for Polish *nie potrzebować*[7] 'to not need' and he claims that the Polish deontic *nie potrzebować* did not prevail completely, i.e., it remains hardly acceptable ("schwankende Akzeptanz") for some Polish native speakers, despite the recommendation of this construction in a prestigious dictionary by Doroszewski (1996).[8]

---

7   Cf. an example given by him, which is only acceptable with *nie musi* 'he does not have to' instead of *nie potrzebuje* 'he does not need to': *On wcale *nie potrzebuje być chory. Być może, nie przychodzi do pracy z innych powodów* 'He does not necessarily have to be ill. It is possible that he does not go to work because of other reasons' (Hansen 2001, p. 150; English translation F.M.).

8   He offers the following example: *Rozmawiamy prywatnie, nie potrzebujesz trzaskać obcasami.* 'We are speaking privately, you do not have to click your heels' (Hansen

In Czech, the situation seems to be quite similar to Polish. The deontic verb *nepotřebovat* 'to not need' is strongly marked. Even though I am not able to present adequate empirical data here, I dare to make a preliminary statement that *nepotřebovat* is peripheral, becoming obsolescent and extinct. This statement can be supported by rapidly decreasing evidence of this construction in Czech dictionaries from the 20th century in contrast to its frequent appearance in the middle of the 19th century, as will be shown below. I propose the explanation that this has resulted from the loss of immediate German-Czech language contact which had helped to sustain this construction.

Before the evidence of the diminishing frequency of the construction, attention should be paid to the criticism of the verb *potřebovati* 'to need' by Czech purists. Surprisingly, in authoritative Czech language handbooks from the last third of the 19th century, only the following usages of *potřebovati* in *affirmative* statements are criticised. Instead of those, simpler verbal forms and a construction with the modal adverb *třeba* 'necessary' are asserted:

> "*Potřebovati*. Zhusta se užívá toho slova ve smyslu něm. *brauchen* chybně, ku př.: 'Potřeboval jsem na něj jen mrknouti,' místo prostého: *jen jsem na něj mrknul*, neb: *bylo mi jen na něj mrknouti*. Du brauchst an den Hund nur zu pfeifen, winken atd.; nikoli: potřebuješ na toho psa jen hvízdnouti, kyvnouti [sic] atd., nýbrž: na toho psa jen *hvízdni, kyvni* [sic] atd. Často také stačí slovo *třeba s infin*.: Třeba jen na něho mrknouti. Třeba mu jen pokynouti atd." (*Matiční brus*, 1st issue 1877, p. 131; similar in the next two issues from 1881 and 1894)[9]

In the mid-19th century, at a time when the German-Czech language contact was much more intensive, but much less reflected in areas other than those concerning direct loanwords or lexical calques, the construction under consideration may have been more frequent. The following data taken from the corpus of Karel Havlíček's correspondence (*KH-dopisy*) seems to support this assumption. In a corpus of only 622K tokens, one can find as many as ten examples of this construction with six different verbs (*psát* 'to write' appears three times, *dělat* 'to

---

   2001, p. 150; English translation F.M.). Furthermore, he proves the restriction of this construction with his ascertainment that it is almost impossible to find *nie potrzebowac* 'to not need' with a non-human first argument, i.e., subject (Hansen 2001).

9 [Czech verb] *Potřebovati*. This word is often used incorrectly in the sense of German *brauchen* 'to need', e.g., 'I only needed to wink at him' instead of the simple 'I only winked at him' or [literally:] 'I could only wink at him'. [German] 'You only need to whistle, wink at the dog etc.': not [the same construction in Czech], but 'whistle/wink at the dog only'. Often it is sufficient to use the word *třeba* 'necessary' with an infinitive: '[It is] necessary to wink at him only/to beckon to him'.

do' and *doložit* 'to prove, to support [a statement]' two times each) as can be seen from the following examples:

(5)   a.   Že jsem přišel domu, *nepotřebuji* Vám *psát* […].
           'I do not need to write to you that I came home.'
           (1841-07-29, František Mudra)

      b.   *nepotřebuješ nic dělát než* všecko s mou adresou dát do […] domu
           'you do not need to do anything else other than to put all things with
           my address into the house'
           (1845-10-02, Vilém Gabler)

      c.   strany budoucnosti *nepotřebuju mít* docela žádnou starost
           'besides the future, I do not need to worry [lit.: "to have any worry"]
           about anything'
           (1847-12-27, Karel Havlíček)

      d.   Milý bratře! Tvoje psaní od 6/1 dostal jsem dnes v poledne a
           nepotřebuji ani *doložit*, že mne tuze potěšilo.
           'Dear brother! I got your letter from January 6th today at noon and
           I do not even need to confirm that it pleased me a lot.'
           (1852-01-12/14, Karel Havlíček)

      e.   štafetu *nepotřebujeme* žádnou *posílat*, všechno se zařídí jednoduchým
           psaním
           'we do not need to send a fast message, everything will be arranged by
           a simple letter'
           (1854-01-24/02-10, Karel Havlíček)

      f.   [včelař] *nepotřeboval by* [včelám] nikdy med *přidávat*, leč v tuze zlý
           rok
           '[a beekeeper] would never need to supplement honey [to the bees],
           only in a very bad year'
           (1854-01-24/02-10, Karel Havlíček)

As written above, an argument for the decline of this construction can be based on the decreasing amount of evidence in 20th-century Czech dictionaries: In the PSJČ (1935–1957), one may find four affirmative and six negative usage examples excerpted from 19th-century authors. The first ones confirm that

the affirmative construction was in use, i.e., the criticism by the purist was, in a sense, "relevant". The second ones should be quoted here:

(6) a. Matice Česká *nepotřebuje si* na špatné účastenství *stěžovati*.
         'Matice česká[10] does not need to moan about bad participation.'
         (Karel Havlíček, 1821–1856)

     b. To jsem už slyšel… *Nepotřebujete* to *opakovat*.
         'I have already heard it. You do not need to repeat it.'
         (Jakub Arbes, 1840–1914)

     c. Dál už *nepotřebuji povídat*.
         'I do not need to talk anymore.'
         (Vítězslav Hálek, 1835–1874)

     d. […] vy mi *nepotřebujete dělat kázaní*, na to je kněz.
         'You do not need to give me a lecture, that's what the priest is for.'
         (Jan Herben, 1857–1936)

     e. *Nepotřebuju nic slyšet*.
         'I do not need to hear anything.'
         (Alois Jirásek, 1851–1930)

     f. *Nepotřeboval* než na tkanici [perly] *navlíkat*.
         'He did not need [to do] anything else other than string pearls.'
         (Karel Jaromír Erben, 1811–1870)

In the SSJČ (1960–1971), only a single example out of the six mentioned, the first one, has survived. Generally, the number of examples in the SSJČ was considerably reduced in comparison to the number given by the PSJČ. The almost complete elimination of negative usage examples, however, indicates a decline of this construction, especially because three of the four examples are of the affirmative *potřebovati* + infinitive left variant:

---

10 Matice Česká, established in 1831, is a Czech cultural institution which functioned as an important foundation during the Czech National Revival.

(7) a. potřebuje si odpočinout
         '(s)he needs to relax'

     b. potřebuje být sama
         'she needs to be alone'

     c. potřebuje jen mrknout, a já už vím, co chce[11]
         '(s)he only needs to wink, and I already know what (s)he wants'

In the SSČ (1978 and following editions), the negated verb *nepotřebovat* in the meaning 'do not need to' is not attested at all. However, the only example of usage which contains the affirmative *potřebovat* + infinitive does deserve attention:

(8) ta věc *potřebuje* ještě pořádně *uvážit*
     'the matter still needs proper consideration'

Instead of this example with the infinitive *uvážit* 'to think about, to consider', a deverbal substantive *uvážení* can be found in the previous dictionary (SSJČ 1960–1971, s.v. *potřebovati*): *věc potřebuje důkladného uvážení*. With a great certainty, its replacement can be explained as a late influence of purism, because deverbal nouns formed from the passive participle were prohibited by many Czech purists (cf., e.g., Jelínek 2007, p. 549–551; Berger, s. a., p. 7). Of particular interest is the fact that a strictly refused Germanism was replaced by another clear one.

Both sections (3) and (4) have documented the potential loss of a modal verb, or a modal meaning of a polysemous (polyfunctional) verb, respectively. In both cases, it would be quite desirable to illustrate the usage of the constructions in question, from the 19th century up to the present, in detail and ascertain (or disprove) their dependence on German-Czech language contact.

5 *Hodlat* 'Intend' and *Uspět* 'Succeed': the Contrasting Development of Two Verbs[12]

So far, we have looked at incomplete processes. The following processes have no clear motivation in German, but they certainly prove that a loss of a modal

---

11 This example is complemented by the following semantic explanation: "stačí, když jen mrkne" 'it suffices/it is sufficient when (s)he winks'.

12 This section is a shortened and revised version of my talk at the conference "Slavic Corpus Linguistics: The Historical Dimension" held in Tromsø (Martínek 2015).

meaning, a process outlined in (3) and (4), can take place and this change does not have to last for centuries. This will be illustrated by two opposite processes in this section.

5.1 *Hodlat* 'intend'

Let us pay attention to the verb *hodlat(i)* first. The earlier form of this verb is *hodlovati*. In Old Czech it is rarely attested. Its original meaning, attested in Tomáš Štítný's writings from the 2nd half of the 14th century, is 'to adjust sth.'—in connection to a concrete, material object. It designates a specific "constructional activity".[13] In another attested context, which is more recent, by Řehoř Hrubý of Jelení (about 1500), the word underwent a semantic change to 'to tailor sth., to make sth. appropriate' and it is joined by an abstract object. Both these meanings are related to the adjective *vhodný* 'appropriate' and to older meanings of the adjective *hodný* (today 'good, kind').[14]

The shift in the meaning of the verb *hodlati* (this form is first attested in 1514) is related to the change in the meaning of the adverb *hodně* 'much, a lot'. In Middle Czech, *hodně* is defined as *náležitě, slušně* 'appropriate' by Veleslavín (which later remains as *vhodně*). This qualitative meaning later undergoes a change and a new, quantitative signification arises meaning 'a lot of, many'. In the first phase, the verb broadens its meaning from designating a specific "constructional activity" to the more general meaning 'to prepare'. This semantic development (bleaching) is reflected by wider collocability—instead of concrete nouns, the verb may later be combined with abstract ones. This induces a subsequent change in the word's meaning: still complemented with an abstract noun, the verb obtains the volitive meaning 'to intend (to do sth., abstract)', 'to want (to do sth.)'. In the later grammaticalization phase, which took place up until the 19th century, the bleached light verb becomes a modal, it stabilizes in this function and its collocability radically changes: now it collocates with verbal infinitives only.

---

13 Note that the subject (agent) is a mason; only the context is figurative.—The term *constructional activity* (*konstrukční činnost* in Czech) was used by Němec for many Old Czech verbs, mostly denominatives, first specialized for designating a concrete activity, later generalized to an action verb (see for example Němec 1987).

14 In contrast, the synonymous Old Czech verbs *(při)hotovati/(při)hotoviti* 'to adjust, to prepare sth.' were able to combine with an infinitive.

In the *19ᵗʰ-century Archive*, one can find 37 suspicious examples among the 600 hits for *hodlati*.[15] In 16 of them, *hodlati* is complemented with a directional adverbial. This can be explained as a lexicalized ellipsis of the motion verb:

(9)   A *kam* ty letos *hodláš*? Nenavštívíš Prahu?
       'Where do you want [to go] this year? Will you not visit Prague?'
       (Božena Němcová, 1820–1862)

In 21 examples, *hodlati* 'intend' occurs without the infinitive verb:

(10)   a.   *Co hodláš*, královno?
              'What do you intend to do, my queen?'
              (Josef Wenzig, 1807–1876)

        b.   Zvěděla také, *co hodlá*.
              'She came to know what he wanted/intended.'
              (Alois Jirásek, 1851–1930)

Two additional remarks on this explanation are needed: It is usually not easy to identify coinages in the dictionaries and distinguish them from regularly used words, and meanings, respectively (e.g, the regional Moravian meaning of *hodlati* 'to laze around', attested by Jungmann and Kott and arising probably due to the influence of the paronymic verb *hovět si*, is not explained here). In other words, the question arises as to whether or not only an analysis of the language centre, the Czech standard and some well-known "near-standard-varieties" is provided here.

Another problem is that dictionaries were based on other dictionaries: some examples appear again in newer dictionaries, although they were probably archaic, functionally and/or stylistically limited, etc. already in the source dictionary.

The verb *hodlati* was finally included in the class of modal verbs; it is not among the *basic modal verbs*, but among the *modal verbs in a wider sense* (see section 1.1). These synonyms of the basic modal verbs can have an additional shade of meaning; and according to Czech dictionaries, it is not easy to decide whether this verb belongs to the central level of the standard language; in some dictionaries, it is assessed as obsolescent and outdated.

---

15   Note that some texts may repeat or be quoted etc. in the excerpts.

5.2 *Uspět* 'to succeed'

The opposing process, which may be called degrammaticalization, can be illustrated by the verb *uspět*, 'to succeed in sth.' in today's meaning. Formerly it also meant 'to achieve, to manage sth.' and '[to manage] to flee, to escape'. This verb can be considered one of the modals—the modal verbs in wider sense, see section 1.1—in Czech of the 19th and the first half of the 20th century. However, during the 20th century, it lost its ability to collocate with the infinitive and its scope of meaning has been reduced to a single one, 'to succeed' ('to fail' in negation, respectively).

This verb is first attested in Jungmann's dictionary (1834-39), although more forms derived from this stem with similar meaning, like *prospěti* 'to benefit sb., to be good for sb.', already occurred in Old Czech. The verb *uspěti* and the substantive *úspěch* 'success', however, are not attested in Czech until the first half of the 19th century.[16] In an etymological dictionary by the Czech linguist Jiří Rejzek (2001), the verb *uspět(i)* is classified as a loanword from Eastern or Southern Slavic languages.

In the case of the substantive *úspěch*, 19th-century dictionaries (Kott's, among others) note polysemy. Kott (1878–1893) defines the first meaning as "uspíšení, die Eile' 'hurry' (cf. *spěch* 'hurry') and the second one as "pokrok, zdar, prospěch, zisk, der Fortschritt, Erfolg". The first meaning refers to a temporal process, the second one primarily to a result. But in fact, all the examples given by Kott exclusively fit the second meaning. The corpus data does not prove this polysemy either. In 1100 example sentences from the *19th-Century Archive*, the substantive *úspěch* is attested only in the meaning 'success'.

As for the verb *uspěti*, the situation is complicated because both these meanings are attested. Its unstable position in the system due to the late loan and the influence of the substantive *úspěch* 'success' probably explain its semantic and functional shift. In the *19th-century Lexical Archive*, *uspěti* occurs twice as a modal verb (see the quoted examples) and 18 times as a non-modal verb.

(11)   a.   Co jsem doposud *vyzkoumati uspěl*, […].
            'What I was heretofore able to find out […].'
            (*Český lid* 1894)

---

16   However, one can find a very suspicious example of the derivate *úspěšnost* 'successfulness' in today's Czech in the *Lexical Database of Humanistic and Baroque Czech*, dated to 1755: *[tkadlec] úspěšnost v díle zadržuje* '[a weaver] delays the success of the work'.

b. A přece *neuspěl jsem probudit* tu sílu Bohem danou ze spánku
'Thus, I did not succeed in waking up the God-given power from
sleeping!'
(Zeyer)

The *Lexical Archive of the PSJČ* contains a huge amount of modal verb usage.
In fiction, it is attested up until the 1960s:

(12) Rozprodala […] pozemky na Žižkově, které nebožtík […] *neuspěl
zastavět*.
'She sold off the pieces of land in Žižkov, which the deceased was not
able to build on.'
(1961)

Also, usage that is uncommon today, without agent as subject, is documented here:

(13) Plány neuspěly.
'His plans did not come true.'
(1956)

5.3 Conclusions

What one can observe here are small, subtle changes in verbal collocability
(semantic valency). The verb *hodlat* 'intend' lost its (fully-lexical) meanings and
became grammaticalized as a modal verb. In this sense, it contrasts with its neu-
tral synonym *chtít*. What are the reasons (motivations) for its use, instead of a
neutral, unmarked synonym? The verb *hodlati* can probably be called "stylisti-
cally higher", i.e., specific for more formal texts—for newspaper style and offi-
cial (administrative) correspondence. By using the 2nd and 3rd person, a distance
from the intention of another person and his/her intention can be expressed, or
the intention can be assessed with light irony. As a (relatively) peripheral lex-
ical unit it can be a source of expressivity in the text as well. The change of the
other verb, *uspět* 'succeed', consists in the reduction of its collocability—in other
words, in the decline of its function as a modal verb. An interesting fact is that it
appears as a modal verb up until the 1960s.

6 *Možná* 'Maybe': The Formation of a Modal Adverb[17,18]

Czech *možná* is a fossilized adjective form, coming from the syntagma *možná věc* 'a possible thing' used as a predicate in connection with the respective form of the verb *být* 'to be'. Due to this circumstance, the rather unexpected feminine form *možná* can be explained, which occurs instead of the—rather expected—neutral form *možné* (or rather parallel to it, cf. *to není možné* 'it is not possible').[19] The modal adverb *možná* 'maybe' emerged as the verb *být* 'to be' was left out.

In Josef Jungmann's dictionary (1834–1839), the word *možná* as a modal adverb 'maybe' is not attested. The first evidence for this form can be found in the corpus *Diakorp*. The sentence in question stems from Karel Hynek Mácha (1810–1836) and was written in 1834/1835.

(14) *Dost možná, že tomu tak bylo.*
      'Quite possible that it was so.'

Based on examples as in (15) from the corpus *KH-dopisy*, which are one to two decades younger than the quoted example by Mácha, several tendencies typical for an ongoing lexicalization process can be documented:

a) In the case of Mácha, možná had not yet become independent. The conjunction *že* indicates that *možná* is still rather seen as an incomplete sentence (as a fossilized predicate, in other words). One could even speculate about *možná že* as one lexical unit but as it is obvious from the examples below, *že* does not have to immediately follow *možná*.

---

17 The development stages of the "epistemic sentence adverb" *možná* can be further characterized using the theoretical background in the paper by Hansen (2010), in which the author considers the analogous development of the Russian *может* (*možet*).

    After elaborating the talk upon which this paper is based, I became familiar with the MA thesis by Fialová regarding the development of Czech modal expressions (Fialová 2013). (My hearty thanks to Hansen for drawing my attention to it.) Fialová (2013, pp. 48–59) describes the rise of the Czech *možná* in detail and focuses on the earlier stages of this process. However, she had a lower number of usage examples, since this paper uses a newer version of the corpus *Diakorp* and in addition, works with the corpus *KH-dopisy* and with the *Lexical Archive of the PSJČ*. Thus, it focuses on the final stage of the development; when it is possible to date the changes more precisely.

18 This section is a shortened part of the talk by me and M. Rybová at the conference "Corpus Linguistics Prague 2016" (Martínek/Rybová 2016).

19 The archaic expression *není možná* '(it is) not possible, no way' currently exists as a gnomic clause without subject coordination.

b) The meaning of *možná* can be modified by other modal words, i.e., strengthened by *dost* 'enough, quite', and weakened by *snad* 'perhaps'. This point also takes the word order into account which is not stabilized yet as one can see in the variation of *dost možná/možná dost*.

c) Apart from this, the expression *(dost) možná* can be coordinated with other modal words, as one can see by the last of the following quoted examples. In this case it is combined with an early stage of the compound which has the form *pravděpodobně* 'probably, likely' (*pravdě* is a dative of *pravda* 'true'; *podobně* 'similar') in contemporary Czech.

(15)   a.   *možná dost že* někdy budeš moci přispěti k vyčistění toho ovčince
             Kristova, kde je hnoje plno
             '[it is] quite possible that you will be able to contribute to the
             cleansing of Christ's sheepfold, which is full of dung'
             (1846-02-23, František Mudra)

       b.   *Možná* až přijdu do Prahy *že* Ti ji povím.
             '[It is] possible that I will tell it to you when I come to Prague.'
             (1846-01-15, František Mudra)

       c.   Nevím, kam se s tím dostanou, *možná že* fortuna slepé povede
             jako posud, *možná* také *že* se zde ve Vlaších dožijem neporádů
             podobných jako někdy we Francii.
             'I do not know where they will get with it, probably [lit.: possible
             that] fortune will lead the blind like it has up to now, probably
             [lit.: possible also that] I will live to see similar disorders here in Italy
             like formerly in France.'
             (1847-03-12/13, František Ladislav Rieger)

       d.   Ostatně buďte ubezpečen: prvé že […], a za druhé že […]; za třetí
             možná ale, pakli se nesmíříte a nás kyselou takovou stravou častovati
             nepřestanete, *že* uveřejním, co jsem psal!
             'After all, be sure that firstly […], secondly […], thirdly, if you will
             not calm down and cease to serve us such sour food, it is possible
             that I will publish what I have written!'
             (1846-06-01, Josef Havelka)

e. *Možná že* by snad to vědomí, žes v nesnázi, mne přece přimělo do něčeho se pustit.
'It may be possible that the knowledge of your troubles would still force me to start something.'
(1846-01-15, František Mudra)

f. *Dost možná a právě podobno, že* ti přece bude moci nějak pomoci […].
'It is quite probable and likely that he would be able to help you somehow.'
(1853-03-01, Karel Havlíček)

In the following examples from *Diakorp*, one can observe that the form *možná* has already become independent from the conjunction *že* one more decade later. However, one has to keep in mind that there is much more evidence of the original usage of *možná že* from this time and that this construction has remained in Czech up to the present.

(16) a. očekávala od něho útěchy [genitive!] a *možná* i nějakou zprávu [accusative!] o Václavovi
'she expected consolation from him and probably also some news about Václav'
(Gustav Pfleger Moravský, *Dva umělci*, 1858)

b. […] *možná* tvá velebnost tomu nejlépe rozumí a já bez rozpaků se podrobuji výroku tvému, otče velebný. [direct speech]
'Your Reverence probably understands it best and I have no qualms about submitting to your verdict, my Very Reverend Father.'
(Prokop Chocholoušek, *Jih*, 1863)

c. "*Možná*," vece chladně vojvoda, "zločin tvůj je všeho příčinou […]."
'"Probably", the duke says calmly, "your crime is the cause of all that".'
(Prokop Chocholoušek, Jih, 1863)

d. […] neboť prý jsou černé oči nejnebezpečnější. (*Možná*, ale čemu?)
'because black eyes are the most dangerous ones. (Maybe, but to what?)'
(S. B. Heller, *Život na Rusi*, 1869)

Based on these text examples, we are able to date the final part of the observed change quite reliably to the 1850s. The last four examples also demonstrate the variability of *možná* which can relate to clause constituents as well as to the whole clause. Apart from functional similarities of the German *möglich* 'possibly' and Czech *možná*, the described development of the Czech modal word does not regard the German-Czech language contact directly. Anyway, in this case it is possible to observe a sequence of "small steps" which lead to language change.

7 Czech *také* and German *auch*

The last empirical section of this paper demonstrates how, similarly to its German counterpart, the Czech modal confirmative particle *také/taky* (German *auch*, English *too*) serves to verify a statement (Cz. *dotvrzovací částice*). Concerning the transition from the original additive meaning (cf. Štěpánková 2014, p. 59) of the adverb Czech *také*, German *auch* 'too', one can find convincing parallels in the development in Czech, German and English. In an overview of the development of English *too*, the usual path described is from the text-organizing function of a discourse marker to the pragmatic function of a particle. This is also true for its German and Czech counterparts:

> We argue that adversative properties of the dialogual discourse context […] appear to have led hearers to reanalyse *too* as expressing a new, rhetorically-strategic meaning with strong counter-argumentative force. The trajectory of change thereby produces a clear path from the ideational/textual meaning of additive *too* to the more clearly interactionally-bound interpersonal meanings associated with non-additive *too*. (Schwenter/Waltereit 2010, abstract)

The word *také* as a confirmative particle, used for verifying statements, is already attested in Old Czech. This circumstance may highlight the role of intralingual development and its causes. Unfortunately, the Old Czech Dictionary (ESSČ) does not contain usage examples. The respective entry, authored by Kateřina Voleková, categorizes the meaning in the following way:

> 1. *také* 'too', 2. *stejně, podobně* 'equally, similarly', 3. *zajisté* 'certainly, of course' [emphasis F.M.]

The confirmative function of *také* remains in New Czech. In the dictionary PSJČ (1935–1957), it is characterized by the following words:

> "dodává tvrzení samozřejmosti, odůvodněnosti", 'it adds the meaning of self-evidence, justification to the statement'

In the next dictionary, SSJČ (1960–1971), this definition is repeated and complemented by an additional one:

"vyjadřuje něj[akou] velkou míru", 'expresses a large degree of something'

Under the first definition, the justification of the statement, one can find a group of very different examples like *co má taky dělat* 'well, what is he to do', sarcastic *to je taky nápad* 'that is a [good] idea indeed', or *čeho se taky bát* 'what are we to be afraid of then'. Under the second definition, regarding a large degree, one can find only the following two examples: *bylo to taky cvičení* 'that was really a [bad] exercise'; *bylo to dnes taky parno* 'it was really hot today'. In my opinion, these two examples do not concern a large degree but, more generally, a speaker's (negative) assessment of the acceptability of a phenomenon or the expression of a (negative) attitude toward it. However, a very similar speaker's attitude is signalized in many examples given for the previous definition; cf. the quoted examples *to je taky nápad* and *bylo to taky cvičení*.

In the dictionary SSČ (1978 and following editions), *také* in *co má také dělat?* 'well, what is he to do?' and *to je také nápad!* 'that is an [good] idea indeed!' is simply characterized as a particle of emphasis (Cz. *zdůrazňovací částice*).

I am not fully satisfied with the formulation that *také/taky* adds the meaning of self-evidence, justification to the statement (cf. PSJČ 1935–1957). I argue that the functions of *také* in the examples quoted are different—this is confirmed by the fact that one would have to use different words to paraphrase them—and that it is desirable to describe them in detail. In my view, these meanings may be precisely described by the first, second and third meaning of the German particle *auch* 'too' as found in the German dictionary DUW:[20]

1. „drückt gefühlsmäßige Anteilnahme, Ärger, Verwunderung o. Ä. aus", 'expresses emotional compassion, irritation, surprise and so on': *der ist auch überall dabei* 'well, he is present everywhere'
2. „bekräftigt od. begründet eine vorangegangene Aussage", 'corroborates or justifies a previous statement': *ich gehe jetzt, es ist auch schon spät* 'I am going now, and also, it is already late'
3. „drückt im Fragesatz einen Zweifel, Unsicherheit o. Ä. aus", 'in an interrogative clause, it expresses doubt, uncertainty and so on': *hast du dir das a. überlegt?* 'did you really think about it?'

The first definition expresses feelings, the second one is close to the additive *too*, the third one concerns certainty.

---

20 Translations of the definitions F.M. Together with each definition, one of the given examples is quoted and translated.

This suggestion—to describe the use of a Czech particle using definitions from a German dictionary—can be seen as an anecdotal confirmation of the importance of German for Czech language development. Moreover, such comparison of particle functions opens up a theoretical discussion about polysemy or the vagueness of particles, i.e., about the possibilities of the discrete separation of the particular meanings. Let us note that another kind of vagueness can be seen by modal verbs where some of their usages are ambiguous, cf. (3) and (4).

8 Summary, Results, Outlook

This paper is an overview of several Czech phenomena regarding modality, complemented with considerations of the German influence on Czech. Moreover, phenomena without a clear German influence were observed, due to and as manifestations of general, recurrent language changes. Of course, further, broader research on this domain is of great importance. Let us suggest some topics which may be of interest.

The aim of papers like this one can be, among other things, to help in the compilation of lexicographic entries that express modality in dictionaries. At this point, I would like to emphasize that I do not mean lexical ("material") Germanisms, loanwords, but rather, syntactic and morphological Germanisms as well as changes in the lexicon: calqued meanings of polysemous words. In the more recent Czech dictionaries from the 20th century, there is still a considerable amount of sentence examples full of Germanisms that are already (sometimes very) obsolete today, but that are not supplied with any appropriate stylistic markers. Other phenomena that were criticised by the Czech purists during the last third of the 19th century and the first third of the 20th century because of their supposed, but in any case not definite German origin, are still portrayed in a negative way, even if these phenomena are not perceived as mistakes by a vast majority of contemporary Czech speakers.

Three analyses of Czech dictionaries should be done in the future to prepare a reliable foundation for further research:

- First, an analysis of which language phenomena expressing modality lost the negative label "from/based on German" (*z/podle němčiny*) and when (in what dictionary) did this happen.
- Second, an analysis of which expressions carry marking other than this negative label, either in the form of a warning "not correct" (*nesprávně*), sometimes with an additional revision, or in the form of the stylistic markers *hovorově* "colloquial (but standard)" and *obecně* "non-standard", signalling the substandard nature of the means in question.

- Finally, an analysis of which phenomena vanish from the dictionaries completely—and does this deletion correlate to their loss in contemporary language (or, more precisely, with the shift of the given means to the language periphery). A related question is, of course, the nature of the relationship between the examples in the dictionaries and their actual usage in communication and whether it is possible that these examples have simply been copied from older dictionaries.

It can also be seen that the Czech morphological and syntactic patterns have receded from the German(ized) models. Two processes play a crucial role in that development: First, the influence of *purism*, which is a *deliberate (intentional)* process of recession, and the *loss of the direct contact* with German speakers after 1945, which is an *unintentional* process, on the other hand. Some examples for the second process are given in this paper: In sections 3 and 4, the decline of the verbs *mít* (cf. German *sollte* 'should') in conditional clauses and *nepotřebovat* 'not to need' in its modal meaning German *nicht wichtig sein* 'to not be important' is demonstrated. Both these usages are significantly marked in today's Czech, one can consider them as obsolescent or even obsolete.

Other examples supporting the idea of the recession of Czech from the German influence are of two types. The substandard Czech construction *jít s sebou* 'to go together with sb.', lit. 'to go with himself', for the German prefix verb *mitgehen* (cf. Giger 2007), has undergone a similar development as the above-mentioned modal verb constructions, i.e., recession. The contemporary broader use of words which were prohibited by purists and temporarily used to a lesser extent, but are not considered Germanisms yet, like *každopádně* 'anyway' (German *jedenfalls*; cf. Vycpálek 1917 and even Filipec 1987, among many others), is an example of the opposite. These processes of the recession and spread of words, word meanings and constructions await further research in the field of German-Czech language contact.

References

Dictionaries

Doroszewski, Witold (ed.):. *Słownik języka polskiego. 18th edition*. Warszawa: PWN 1996.

DUW = Klosa, Annette et al. (eds.): *Duden – Deutsches Universalwörterbuch. 4th edition*. Mannheim et al.: Dudenverlag 2001.

ESSČ = *Elektronický slovník staré češtiny* (online). Praha: Ústav pro jazyk český AV ČR 2006–2018. Retrieved at 21. 12. 2018, from http://vokabular.ujc.cas.cz.

Jungmann, Josef. *Slovník česko-německý*. Praha: Knížecí arcibiskupská knihtiskárna 1834-1839. Retrieved at 21. 12. 2018, from http://vokabular.ujc.cas.cz/moduly/slovniky.

Kott, František Štěpán: *Česko-německý slovník zvláště gramaticko-frazeologický*. Praha: František Šimáček 1878-1893. Retrieved at 21. 12. 2018, from http://kott.ujc.cas.cz/.

Lexikální databáze humanistické a barokní češtiny (online). Praha: Ústav pro jazyk český AV ČR 2010–2018. Retrieved at 21. 12. 2018, from http://madla.ujc.cas.cz.

PSJČ = *Příruční slovník jazyka českého*. Praha: Česká akademie věd a umění/ Státní pedagogické nakladatelství 1935–1957. Retrieved at 21. 12. 2018, from http://psjc.ujc.cas.cz, including its Lexical Archive.

Rejzek, Jiří: *Český etymologický slovník*. Praha/Voznice: Leda 2001.

SSČ = *Slovník spisovné češtiny*. Praha: Academia 1978 (and following editions).

SSJČ = *Slovník spisovného jazyka českého*. Praha: Academia 1960–1971. Retrieved at 21. 12. 2018, from http://ssjc.ujc.cas.cz.

Corpora

Diakorp: Kučera, Karel, Anna Řehořková & Martin Stluka. 2015. *Diakorp*. Version 6 [4M tokens]. Praha. Retrieved at 21. 12. 2018, from http://www.korpus.cz

KH-dopisy: Adam, Robert et al. 2017. *Korpus korespondence Karla Havlíčka* [622K tokens]. Praha. Retrieved at 21. 12. 2018, from http://www.korpus.cz.

19th-Century Archive: Non-published corpus of Czech 19th-century texts, available in the Institute of the Czech National Corpus, Prague.

Staročeská textová banka. Praha. Retrieved at 21. 12. 2018, from http://vokabular.ujc.cas.cz/banka.aspx?idz=STB.

Scientific literature

Berger, Tilman: "The Influence of German on the Grammatical System of Czech". In: Spieß, Gunter (ed.): *Deutsch-slawischer Sprachkontakt. Entlehnungen und sprachliche Identität*. Frankfurt/Main: Lang (to appear).

Berger, Tilman: „Deutsche Einflüsse auf das grammatische System des Tschechischen". In: Berger, Tilman. *Studien zur historischen Grammatik des Tschechischen. Bohemistische Beiträge zur Kontaktlinguistik*. München: Lincom 2008, pp. 57–69 (The preprint version of this article (24 pp.) is available on: https://homepages.uni-tuebingen.de//tilman.berger/Publikationen/BergerCottbus.pdf.).

Ertl, Václav: „K smrti, k jídlu, k pláči, k dostání a pod". In: *Naše řeč* 11 (10), 1927, pp. 217–229.

Fialová, Markéta: *Zum Wandel der unpersönlichen Modalausdrücke možn-, lze und jde im Tschechischen. Eine korpusbasierte Studie zum Mikrowandel*. University of Regensburg: MA-thesis 2013.

Filipec, Josef: „Čeština z obrazovky". In: *Naše řeč* 70(1), 1987, pp. 10–21.

Fried, Miriam: "The polyfunctional structure of *jestli* in spoken Czech". In: Štícha, František/Šimandl, Josef (eds.): *Grammar and Corpora I*. Praha: Ústav pro jazyk český AV ČR 2007, pp. 57–67.

Giger, Markus: „Jít s sebou a podobná spojení v ČNK a na internetu". In: *Naše řeč* 90(4), 2007, pp. 195–202.

Grepl, Miroslav: „Vyjadřování modalitní kategorie záměru". In: *Slovo a slovesnost* 34(1), 1973, pp. 74–77.

Hansen, Björn: *Das Modalauxiliar im Slavischen. Semantik und Grammatikalisierung im Russischen, Polnischen, Serbischen/Kroatischen und Altkirchenslavischen*. München: Sagner 2001.

Hansen, Björn: "Constructional Aspects of the Rise of Epistemic Sentence Adverbs in Russian". In: Hansen, Björn/Grković-Major, Jasmina (eds.): *Diachronic Slavonic Syntax. Gradual Changes in Focus: Wiener Slawistischer Almanach, Sonderband 74*. München/Berlin/Wien: Kubon & Sagner 2010, pp. 75–86.

Hansen, Björn/Nekula, Marek/Banášová, Monika: "Nová konstrukce 'Karla Gotta nemusím' v češtině a slovenštině: případ lexikalizace, pragmatikalizace nebo začínající degramatikalizace?" In: *Slovo a slovesnost* 72(4), 2011, pp. 243–267.

Jelínek, Milan: "Purismus". In: Pleskalová, Jana (ed.). *Kapitoly z dějin české lingvistické bohemistiky*. Praha: Nakladatelství Lidové noviny 2007, pp. 540–573.

Karlík, Petr/Štícha, František: "Infinitivní fráze v některých typech syntaktických struktur". In: Štícha, František (ed.): *Kapitoly z české gramatiky*. Praha: Academia 2011, pp. 928–944.

Kolářová, Ivana: "Některé významy a funkce sloves podívat se, koukat/ kouknout se". In: *Naše řeč* 82(2), 1999, pp.65–71.

König, Werner/Elspaß, Stephan/Möller, Robert/Paul, Hans-Joachim: *dtv-Atlas Deutsche Sprache*. München: Deutscher Taschenbuchverlag 2015.

Martínek, František: *Peripheral Czech modal verb + infinitive constructions*. Paper presented at the conference "Slavic Corpus Linguistics: The Historical Dimension" at The Arctic University of Norway, Tromsø, April 2015.

Retrieved at 21. 12. 2018, from http://site.uit.no/slavhistcorp/files/2015/08/
Martinek.pdf.

Martínek, František/Rybová, Martina: *Modální částice v korespondenci Karla
Havlíčka*. Paper presented at the conference "Corpus Linguistics Prague
2016" at the Charles University, Prague, September 2016. Retrieved at
21. 12. 2018, from http://ucnk.korpus.cz/kl2016/files/slideshow/166.pdf.

Matiční brus: *Brus jazyka českého, jejž sestavila komise širším sborem Matice
české zřízená*. 1877. (Next issues 1881 and 1894.) Praha. Retrieved at
21. 12. 2018, from http://vokabular-test.ujc.cas.cz/moduly/mluvnice/
digitalni-kopie-detail/Brus1877.

Nekula. Marek: *System der Partikeln im Deutschen und Tschechischen.
Unter besonderer Berücksichtigung der Abtönungspartikeln*. Tübingen:
Niemeyer 1996.

Němec, Igor: "Předložka *po* se substantivem jako základ sloves a jejich
etymologické východisko". In: *Naše řeč* 70(4), 1987, pp. 177–184.

Newerkla, Stefan Michael: "Areály jazykového kontaktu ve střední Evropě a
německo-český mikroareál ve východním Rakousku". In: *Slovo a slovesnost*
68(4), 2007, pp. 271–286.

PMČ = Grepl, Miroslav/Karlík, Petr et al. (eds.): *Příruční mluvnice češtiny*.
Praha: Nakladatelství Lidové noviny 1995.

Schwenter, Scott/Waltereit, Richard: "Presupposition accommodation and
language change". In: Davidse, Kristin/Vandelanotte, Lieven/Cuyckens,
Hubert (eds.): *Subjectification, Intersubjectification and Grammaticalization*.
Berlin/New York: Mouton de Gruyter 2010, pp. 75–102.

Štěpánková, Barbora: *Aktualizátory ve výstavbě textu, zejména z pohledu
aktuálního členění*. Praha: Ústav formální a aplikované lingvistiky MFF
UK 2014.

Thomason, Sarah Grey/Kaufman, Terrence: *Language contact, creolization, and
genetic linguistics*. Berkeley: University of California Press 1988.

Vycpálek, Josef: "Pád". In: *Naše řeč* 1(5), 1917, pp. 151–152.

Weiss, Daniel: "Polsko-niemieckie paralele w zakresie czasowników
modalnych (na tle innych języków zachodniosłowiańskich)". In: Hentschel,
Gerd/Ineichen, Gustav/Pohl, Alek (eds.): *Sprach- und Kulturkontakte im
Polnischen*. München: Sagner 1987, pp. 131–156.

Weiss, Daniel: „Mögliche Argumentationen zum Nachweis von Calques
am Beispiel der polnischen Modalverben". In: Scholze, Lenka/Wiemer,
Björn (eds.): *Von Zuständen, Dynamik und Veränderung bei Pygmäen und
Giganten*. Bochum: Universitätsverlag Dr. N. Brockmeyer 2009, pp. 129–153.

Stefan Michael Newerkla

Linguistic Areas in East-Central Europe as the Result of Pluridimensional, Polycentric Convergence Phenomena

Abstract: The following contribution deals with problems that make analysing linguistic areas in East-Central Europe, as the result of pluridimensional, polycentric convergence phenomena, a real challenge. It demonstrates why we still must investigate further into the contribution of specific groups of people to the emergence of certain areas in order to gain a better understanding of linguistic areas, especially in East-Central Europe. Furthermore, it also shows that in this context it seems more appropriate to speak of polycentric rather than pluricentric convergence. Whereas a pluricentric language is the sum of its varieties, a polycentric language according to Li/Juffermans (2012, p. 77) is "a dynamic, socially ordered system of resources and norms that are strongly or weakly associated with one or more centers". As we could see from the example of their different evolution and history, signed languages are not so tied to the spoken languages of a region, but rather to a place or a social stratum. This fact makes the concept of linguistic areas appear even more vivid and dependent on social interaction rather than on the specific characteristics of the languages in contact: Languages do not converge by themselves, it is the behaviour of the speakers that brings about these pluridimensional, polycentric convergence phenomena leading to specific linguistic areas.

Keywords: linguistic areas, East-Central Europe, linguistic convergence, polycentricity, historical sociolinguistics

1 Introduction

For more than a century, linguists from different cultural backgrounds have been using the term "linguistic area" to denote languages that have developed common features resulting from geographical proximity and language contact. Rik van Gijn and Pieter Muysken (2016, s. p.) define these areas "as social spaces (regions, countries, (sub-)continents) in which languages from different families have influenced each other significantly, leading to striking or remarkable structural resemblances across genealogical boundaries." Despite that, as Sarah Grey Thomason (2000, p. 311) aptly remarks, there is still little consensus on the general nature of the phenomenon, although there are numerous valuable studies of particular linguistic areas and of particular features within certain linguistic areas. This is certainly caused by the complexity of the situation, or as Thomason

(2000, p. 311) puts it, "The most important (though not very neat) conclusion, however, is that attempts to find very general social and/or linguistic principles of convergence in a linguistic area are doomed – not only because every Sprachbund differs from every other one, but also because the conditions of contact in large Sprachbünde will inevitably vary over time and space." In other words, areas of linguistic convergence are diffusion areas or varying language crossroads and thus not a uniform linguistic, social or historical phenomenon.

Moreover, since the approaches to the study of the distribution of linguistic features have been mostly structural and historical, the notion of "linguistic areas" has been much criticised in the strict sense. In tandem with a better understanding of the psycholinguistic and sociolinguistic mechanisms and scenarios leading to linguistic areas, the areal perspective keeps gaining ground, again, in explaining how languages actually converge and which mechanisms promote or block this type of convergence: "Languages do not converge by themselves; rather, it is the agency or unconscious behavior of speakers that has this effect." (Gijn/Muysken 2016, para. 1)

2 Linguistic Areas in East-Central Europe

Now, if we have a look at East-Central Europe, we are confronted with various contact areas of Germanic languages with Baltic, Finno-Ugrian and Slavic languages. Roughly since the 6th and 7th centuries, Slavs had settled the lands in Central and Eastern Europe including much of present-day Germany and Austria, abandoned by Germanic tribes fleeing the Huns and their allies. We can find traits of this settlement in many place names east of the line of the Elbe and Saale rivers today. In the following centuries, so-called marches were established east of this line to protect the frontier, from which an eastwards colonisation into Slavic territory commenced. Moreover, the subsequent expansion of the Magyars as well as the Bavarianisation of the region of present-day Austria separated the northern and southern Slavs. However, their influence on the languages of the people in – at least eastern – Austria has remained intact ever since and has become even more manifest in the wake of the major waves of Slavic migration to Vienna in the 19th and 20th centuries (Newerkla 2000, 2012a, 2012b, 2012c, 2013a, 2013b).

At the same time, the large north-south extension of German and its spread over several countries and, subsequently, states has led to the rather uncontroversial conclusion that German is a pluricentric language. Not only does pluricentric German display characteristic features of Standard Average European, but it also comprises several distinguishing features in various contact areas with Baltic,

Finno-Ugrian and Slavic languages. Therefore, it seems justified to speak not just of one East-Central European language area, but of several varyingly distinct and overlapping language contact areas in Eastern Central Europe. Like isoglosses, which constitute certain dialect areas in dialectology, bundled language contact phenomena distinguish certain contact areas from others. In this context, further research on the role of Yiddish for the emergence and understanding of linguistic areas in Eastern Central Europe is still a major and pressing desideratum.

2.1 Pluridimensional Convergence – The Example of Austria

A major language contact area in East-Central Europe – merely one out of several – is the contact zone with the former centre of the Austro-Hungarian Empire, with German, Czech, Slovak and Hungarian as its core languages as well as Polish, Slovene and Bosnian-Croatian-Serbian as its only partially involved peripheral languages. A detailed description of this area and the history of its evolution can be found in Jiří Januška's (2017) new dissertation, comparing Central European languages beyond structural features and loanwords.[1] He is also one of the contributors to this book.

In present-day Austria, we can still identify traces of this multilingual area. There are seven officially recognised minority languages, the languages of the so-called six autochthonous ethnic groups officially recognised by the Ethnic Groups Act (VoGrG): Burgenland Croatian, Slovene, Czech, Slovak, Hungarian and Romani, plus Austrian Sign Language (*Österreichische Gebärdensprache*, ÖGS). However, the 20th century also brought about a significant change in the importance of the several ethnically Slavic minority groups and their languages in Austria. Whereas, for example, the influence of Czech and Slovak declined, the importance of other groups – e.g. the Poles (after 1978), but especially the Serbs, Croats and Bosnians – and their languages increased successively throughout the second half of the 20th century. These immigrant workers arrived in large numbers in the wake of the war in the Balkans and in parallel with the increased Turkish population in Austria.

To date, a considerable amount of literature on Slavic-German language contact phenomena has been published (the relevant chapters in Goebl/Nelde/Starý/Woelck 1996–1997 and the bibliography in Newerkla 2011, pp. 619–710). In this context, one of the most promising efforts to reconcile the fragmented research community on German in East-Central Europe was the launch of the

---

1   A recent important achievement on this topic is also the summarising book on the Central European languages by Ondřej Bláha (2015).

Research Centre for German in Central, Eastern and Southeastern Europe (FZ DiMOS) in 2014 as a scientific institution of the Faculty of Linguistics, Literature and Cultural Studies (SLK) of the University of Regensburg.[2] Its primary goals are analysing and documenting the German language in East-Central Europe by taking into account the historical and current multilingual situation of this area, and by cooperating closely with colleagues from the local universities and other scientific institutions. At present, German no longer takes the role of a dominant language, but functions as an interregional means of communication and as a bridge language in an area stricken by modern migration movements.

In contrast, comparably minimal systematic and exhaustive linguistic research has been conducted on the linguistic influences and contact phenomena between the Slavic languages (including their varieties) and German in Eastern Austria. Recent studies on the matter are rare (the last comprehensive study being Steinhauser 1978), or only highlight certain aspects (e.g. Ernst 2008, Masařík 1998, Newerkla 2007a, 2007b, 2009, Pohl 1999, 2007, Zeman 2009). However, several popular descriptions of these phenomena have been published since the 1980s (e.g. Grüner/Sedlaczek 2003, Schuster/Schikola 1996, Sedlaczek 2007, 2011, Wehle 1980, 1996, 1997). However, some of them partially comprise unverified information and perpetuate language myths.

In 2016, a consortium consisting of Alexandra Lenz, Gerhard Budin and Stefan Michael Newerkla from the University of Vienna, Stephan Elspaß from the University of Salzburg and Arne Ziegler from the University of Graz were granted a Special Research Program (SFB) by the Austrian Science Fund (FWF) on "German in Austria. Variation – Contact – Perception" (F 60-G 23).[3] The scope and topic of this SFB encompass the entire spectrum of variation and varieties of German in Austria, bringing together expertise from the fields of variationist linguistics, contact linguistics and multilingualism research, as well as from sociolinguistically based research on language perception and attitudes. Project part "German and the Slavic languages in Austria. Aspects of language contact" in task cluster C will eventually culminate in a detailed overview of contact-induced Slavic influences on the varieties of German in Austria over time by concentrating on the exemplary situation in the urban area of Vienna. Whereas one part of our research is aimed at the historic dimension of language contact,

---

2 See FZ DiMOS: http://www.uni-regensburg.de/forschung/dimos/ (accessed 13/05/2019).

3 See SFB "German in Austria. Variation – Contact – Perception": http://www.dioe.at (accessed 13/05/2019).

in which Czech was the dominant contact language, the other will address the present-day situation, in which Bosnian-Croatian-Serbian and Polish are the most important Slavic varieties in contact with the German spoken in Vienna. By doing so, we will be able to identify parallels with and contrasts to the former situation. In particular, we want to find comprehensive answers to the following research questions: What was the effect of language contact with Czech and other Slavic languages on the different language levels of the varieties of German in the city and agglomeration area of Vienna, especially during the peak of Vienna's Czech minority in the last decades of the Habsburg Empire? What is the effect of language contact with Slavic languages, especially Serbian, Croatian, Bosnian and Polish, on the individual language levels of the varieties of German in the urban area of Vienna today? In addition, can we identify any comparable, special or universally applicable aspects of language contact in this linguistic area? At this moment, we are still in the process of data collection and analysis, but my co-workers and project members Agnes Kim and Maria Schinko already present partial research results in their contributions to this book. That is why here and now, I just briefly recapitulate and summarise the results of our previous research in the field (especially Newerkla 2007a, 2007b, 2009, 2013a, 2013b). On the one hand, we can identify a clear convergence of the vocabularies of at least Czech, Slovak, Hungarian and German standard as spoken in Old Austria, which I have already touched upon in other papers (Newerkla 2000, 2007a, 2007b, 2009, 2011, pp. 76–86). There are many German loanwords in Czech, Slovak and Hungarian that derive from German words, which are still or at least were in use solely in the Austrian variety of High German. On the other hand, many Slavic, Hungarian and also Romance words found their way into the German of Old Austria and thus set to a great extent the typical character of the Austrian variety of standard High German (e.g. *Buchtel, Klobasse, Zipp, Automatenbuffet, Chauffeur, Fauteuil, Garçonnière, Lavoire, Plafond, Bartwisch, Busserl, Bussi, Dekagramm, Fasching, fesch, Hetz, Semmel, Werkel, Zeller, Biskotten, Karfiol, Malter, sekkieren, Trafik, Adjunkt, Evidenz, lizitieren, Matura, Ribisel, paprizieren, Palatschinke, Pogatsche, Kukuruz,* and so on).[4] Many of them were again passed on to other languages of the Habsburg Empire through the medium of Austrian German.

---

4   The English equivalents are in succession of their appearance: *yeast pastry; hard smoked sausage; zip-fastener; automat; chauffeur; armchair; bed-sitter; wash-basin; ceiling; hand-brush; little kiss; 10 grammes = 154,324 grains* (troy and avoirdupois)*; Shrovetide; smart; fun; (Vienna) roll; barrel-organ; celeriac; biscuits; cauliflower; mortar; pester; tobacco-shop; assistant director* (one of the innumerable titles of civil servants in the

This is in accordance with an observation by Roman Jakobson (1938, p. 52) from the first half of the 20th century. He pointed out the fact that the limits of language convergence seem to coincide strikingly with boundaries of physical and political geography. By stating so, he anticipated later findings of the American sociolinguist Dell Hymes (1974), who claimed that different languages could form a speech community under certain political influence and social conditions.

George Thomas from McMaster University, Hamilton, Ontario investigated the role of German loanwords in the Slavic languages of the Habsburg Empire on a larger scale taking into account Czech, Slovak, Slovene and Croatian. The results of his statistical evaluation among other things clearly show the important integrating function of the Austrian variety of German at that time by providing a list of German loanwords common in all the languages analysed, whereas the individual Slavic equivalents correspond only in 16 % of the instances ascertained. (Thomas 1997, pp. 341–349). We can therefore find many of the most common German loanwords in Czech also in the other languages of Old Austria, especially in their colloquial variants. In this regard, Emil Skála (1998, p. 217) mentions words such as *Gesindel – ksindl – ksindl – kszindli* 'scoundrels, riff-raff' or *Schwindel – švindl – švindl – svindli* 'swindle, cheat'.[5] Certainly, Skála's remark has its validity, but I think he does not really get to the core of the whole thing by unluckily omitting one very important fact, i.e. that the borrowing processes proceeded in several directions and thereby led to many agreements among the distribution of semantic content. As a result, these languages have become semantically similar while remaining phonetically diverse to some extent.

Such processes of language convergence become even more evident, if we do not confine ourselves just to German loanwords, but look at shared linguistic phenomena as such, for example, the use of prepositions in Austrian German, Czech and Slovak as well as the use of the corresponding suffixes in Hungarian. In English and in German as spoken in Germany we take an examination in a subject such as Russian, mathematics and so on (= *eine Prüfung in Russisch, Mathematik, … ablegen*). However, the equivalents in Austrian German, Czech, Slovak and Hungarian are in this succession *eine Prüfung aus Russisch, … ablegen*; *vykonat zkoušku z ruštiny, …*; *vykonať skúšku z ruštiny, …*; *oroszból, … vizsgáz(ni)*. The meaning of the prepositions *aus, z/ze* and *z/zo* as well as the

---

Habsburg Empire)*; register; sell by auction; school-leaving exam; currant; to spice with paprika; pancake; pancake with greaves; Indian corn.*

5   We cite the examples in the following order of languages: German, Czech, Slovak, and Hungarian.

Hungarian suffixes *-ból/-ből* is the same (literally 'out of, from, of'). Similarly, in English and German as spoken in Germany we sit at the table (= *am Tisch sitzen*); the equivalents in Austrian German, Czech, Slovak and Hungarian are *bei Tisch sitzen; sedět u stolu; sedieť pri stole; asztalnál ül(ni)*. The meaning of the prepositions *bei, u* and *pri* as well as the Hungarian suffixes *-nál/-nél* is the same, again (literally 'near, close to'). A striking feature of Austrian German – especially of its colloquial varieties – in contrast to German as spoken in Germany is also the extensive and unmarked use of the preposition *auf* (= literally 'on, upon'): *auf der Universität, auf der Post, auf dem Hof, auf dem Konzert, auf dem Markt*. In many cases, this characteristic can once more be associated with the use of the preposition *na* in Czech *na univerzitě, na poště, na dvoře, na koncertě, na tržišti*, …, and Slovak *na univerzite, na pošte, na dvore, na koncerte, na trhovisku*, …, as well as the use of the Hungarian suffix *-n* (*-on, -en, -ön*) with the same meaning az *egyetemen, a postán, az udvaron, a koncerten, a piacon*, …. (cf. Newerkla 2011, p. 80).

However, lingua-cultural convergence also affected the conceptual world of the urban spaces in the Habsburg Monarchy and subsequently the population throughout the Empire. Among other things, this led to certain brand and product names (known to many people even today, Newerkla 2012c). The company name Pischinger is just one example. Founded by Oscar Pischinger in 1849, it created the famous and still popular *Original Pischinger Torte*, a cake made of special cake-sized round wafers. During its heyday, the Vienna-based company had over 500 employees and outlets in Bratislava, Cracow, Chernivtsi, Budapest and Osijek (cf. Czech *pišingr*, Slovak, Slovene, Bosnian-Croatian-Serbian *pišinger*, Polish *piszinger* and Hungarian *pisinger*). Another example is the brand name Ceres, with later an Austrian and a Czech version of this coconut fat (still being sold). Further products are e.g. the Austrian grape variety *Zierfandler*, Moravian Czech *cinifádl, cinifál*, Slovak *cirifandel, cirifandl, cilifandl*, Hungarian *cirfandli* (earlier *cirifandli, tzirifándli, cilifánt*) not to be mixed up with the *Zinfandel* also known as *Primitivo* (Gold 2009). The *Kaisersemmel* 'Kaiser roll' (also called a 'Vienna roll'), is a typically crusty round bread roll, originally from the Austrian Netherlands. Again, the Kaiser rolls have become popular throughout the Austrian Habsburg Empire. Today, they are also known in Poland (Galicia), the Czech lands, Slovakia, Hungary, Slovenia, Croatia, and even parts of Italy, Germany, the United States, and Canada (Newerkla 2012c). German *Teebutter* and the subsequent translations into Czech *čajové máslo*, Slovak *čajové maslo*, Hungarian *teavaj*, Slovene *čajno maslo*, and Croatian *čajni maslac* arose like many other German compounds with *Tee-* in the wake of the popularisation of English tea customs in East-Central Europe, especially the

tradition of serving tea with – at least – bread and butter. In agreement with the English preferring creamery butter to butter made out of sour cream, Austrian German *Teebutter* and its equivalents in the neighbouring languages originally denoted only best quality creamery butter. Later on, the notion became the official designation of best quality butter. As such, we can find the term in Austria's food code, the Codex alimentarius Austriacus, up to this day (Newerkla 2008).

Language use of this kind – both written and oral – not only reflects social patterns, but also the interrelatedness of discursive practices and cultural encounters. However, the role of transnational linguistic practices in people's everyday lives has so far been rather neglected, although the Habsburg monarchy was clearly a contact zone of migrants and travellers, a linguistic area where people drew on the practices of their various places of origin. From this linguistic area, a micro-area emerged in Vienna and Eastern Austria that was particularly affected by the influence of Czech on German (Newerkla 2007a). As early as the 19th century, the knowledge of Czech loanwords in Vienna was so strong that the well-known Austrian actor, singer and playwright Johann Nepomuk Nestroy could make use of them in his comedies and other dramatic pieces. We identified at least 50 words ranging from *ale* 'but' in his play *Martha oder Die Mischmonder Markt-Mägde-Mietung* (1848) to the pejorative denotation of Czechs as *Zopaks* (derived from *copak* 'what?') in his play *Eisenbahnheiraten oder Wien, Neustadt, Brünn* (1844). Other expressions used by Nestroy and then commonly known were *heidipritsch* 'totally gone' (< onomatopoetical *hajdy* and *pryč* 'gone'), *hubitschko* 'peck on the cheek' (< *hubička*), *Kaluppe* 'dilapidated, ramshackle hut' (< *chalupa* 'hut'), also in the German diminutive form *Kalupperl*; *Leschak* 'lay-about' (< *ležák*), *nemam* 'have-not' (< *nemám*), *petschieren* 'seal' (< *zapečetit*), *powidalen* 'tell' (derived from the preterite form of *povídat*), *Rosimi (-sim-/-sum-)* 'wits' (< *rozum*), etc. (cf. Newerkla 2009, p. 9, 2013a, p. 254).

The influence of Czech and Slovak was also very strong in the semantic field related to cooking (kitchen words, denotations of food and meals). Words such as *Bramburi* 'potatoes' < *brambory*; *Buchtel* (B-/W-) 'yeast pastry' < *buchta*; *Liwanze* 'pancakes' < *lívanec*; *Klobasse* (-e/-i) 'hard smoked sausage' < *klobása*; *Kolatsche* (K-/G-) 'small yeast cake with filling' < *koláč* 'cake'; *Oblate* (stressed on the first syllable as in Czech) 'fine wafer' < *oplatka*; *Palatschinke* < *palačinka* (< Hungarian *palacsinta* < Romanian *plăcintă*) 'jam-filled pancake'; *Powidl* 'plum jam' < *povidla*; *Skubanki* (Sk-/St-) 'sweet noodles with poppy seeds' < *škubánky*; but also *Brimsen* 'sheep's milk cheese' < Slovak *bryndza* (< Romanian *brînză* apart from *brânză* 'cheese'); *Haluschka* 'chopped cabbage fried in butter and served over boiled noodles' < Slovak *haluška*, etc. are commonly known in Vienna even today (cf. Newerkla 2009, p. 9, 2013a, p. 254).

The language contact with Czech also had immediate influence on word formation in the colloquial variety of German in Vienna. This can be seen, e.g. in the use of the Czech word formation suffix *-ák* in words not known in Czech such as *Böhmak* 'Czech male', *Feschak* 'dashing young man', *Tränak* 'camp follower' (< French *train* and *-ák*), etc. But also the use of Czech stems with German word formation suffixes can be found, such as *Tschunkerl* 'mucky pup' < *čuně* 'piglet' and the Bavarian diminutive suffix *-erl*, or mixed suffix forms, such as *Armutschkerl* 'poor wretch' with two combined diminutive suffixes (< Czech *-č(e)k-* and Bavarian *-erl*). Even German verbs could be derived from Czech words, such as *verdobrischen* 'squander, blow' < *dobrý* 'good' (cf. Newerkla 2009, p. 9, 2013a, pp. 254–255).

To this day, we can encounter persons in all spheres of Vienna's public life, whose ancestors were born in the Czech lands and Upper Hungary, or who at least have Czech or Slovak family names. Simply consider the family background of the former Austrian chancellor Bruno Kreisky, or the former Viennese mayor and subsequent president Franz Jonas, or the Czech names of other Austrian politicians such as Blecha < *blecha* 'flea'; Busek < *Bušek*, a diminutive of *Buš* derived from the name *Budimír, Budislav, Budivoj* or *Bohuslav*; Cap < *čáp* 'stork', Ceska < *čieška* 'small bowl' in Old Czech; Dohnal < *dohnal* 'he who caught up with'; Klestil < *klestil* 'he who pruned'; Klima < *Kliment* (Latin *Clemens*); Kukacka < *kukačka* 'cuckoo'; Lacina < *laciný* 'cheap', etc. Some Czech family names have become denotations of certain typical characters, e.g. *Březina*, *Novák* and *Trávníček* in expressions such as *Na servus Brežina!* in order to express unpleasant surprise; *Er ist immer der Nowak* in the sense of 'he is always the victim, he is always abused'. *Trawnitschek* is the embodiment of the typical petty bourgeois, known in Austria as the alter ego of the former actor Helmut Qualtinger (cf. Newerkla 2009, p. 8, 2013a, p. 253).

The code switching from Czech to German has over the centuries led to the characteristic use of prepositions in the Viennese colloquial variety of German. Take for example the equivalent prepositional phrases *auf Urlaub fahren* < *jet na dovolenou* 'go on holiday', vs. Standard German *in Urlaub fahren*; *auf zwei Tage nach Prag fahren* < *jet na dva dny do Prahy* 'travel to Prague for two days', vs. Standard German *für zwei Tage nach Prag fahren*; *auf jmdn./etw. denken* < *myslet na někoho/něco* 'think of someone', vs. Standard German *an jmdn./etw. denken*; *Vorbereitungen auf etw.* < *přípravy na něco* 'preparations for something', vs. Standard German *Vorbereitungen für/zu etw.*; *in der Nacht auf Sonntag* < *v noci na neděli* 'in the night to Sunday', vs. Standard German *in der Nacht zum Sonntag*; *sich auf jmdn./etw. erinnern* < *vzpomenout si na někoho/něco* 'remember someone', vs. Standard German *sich an jmdn./etw. erinnern*; *auf jmdn./etw.*

vergessen < zapomenout na někoho/něco 'forget someone', vs. Standard German *jmdn./etw. vergessen* (cf. Newerkla 2007a, p. 281, 2007b, p. 40).

Czech and the languages of other Slavic immigrants also fostered the use of hypocoristics and diminutives in Viennese German such as *Anči* for Anna or *Mamitschka* for mummy (< *mamička*) as well as the so-called double negation of the type *er hat kein Geld nicht g'habt* as in Czech *neměl žádné peníze* 'he did not have any money', *sie hat niemandem nichts gesagt* as in Czech *nikomu nic neřekla* 'she did not tell anyone', etc. (cf. Newerkla 2009, p. 10, 2013a, pp. 255–256).

Further results of this code switching from Czech to German in Vienna are phrases such as *Er/sie soll sich ausstopfen lassen! < Ať se jde vycpat!* in the sense of *Zum Kuckuck mit ihm/ihr!* 'Damn him/her!'; *Ohne Arbeit gibt's keine Kolatschen! < Bez práce nejsou koláče!* in the meaning of *Ohne Fleiß kein Preis!* 'no pains, no gains'; *die Kinder spielen sich < děti si hrají*, German *die Kinder spielen* 'the children play'; *Sonst bist g'sund? < Jinak si zdravý?* in the sense of *Bist du (noch) bei Trost?* 'Have you gone mad?'; *die Patschen strecken < natáhnout papuče/ bačkory* for German *versterben* 'pass away'; *sich etw. aus dem Finger zuzeln < něco si vycucat z prstu* in the meaning of *etw. erahnen, erfinden* 'make something up'; *es steht (sich) (nicht) dafür < (ne)stojí to zato* in the meaning of *es lohnt sich (nicht)* 'it is (not) worth the effort'; *seine sieben Zwetschken packen < sbalit si svých pět švestek* (in Czech there are just five plums), in the sense of *sein Hab und Gut packen und gehen* 'to pack everything one owns and move to another place'; *das geht sich (nicht) aus < to (ne)vyjde* for German *das klappt (nicht)* 'turn out well/badly, work out all right'; *Das ist nicht mein Gusto! < To není mé gusto!* in the sense of *Das ist nicht mein Geschmack!* 'This is not my liking!', etc. These phrases have been integrated into Austrian German to such an extent, that we no longer perceive them as foreign, but as language elements typical of the Austrian variety of German. Other typically Viennese phrases are e.g. *auf Lepschi gehen* 'enjoy oneself' equivalent to Czech *jít na lepší*; *außer Obligo sein* 'be free of any obligation' < *být z obliga*; *bridsch sein* in the sense of 'be gone, be lost' < *pryč*; *na servus!* meaning 'fancy that' and expressing unpleasant surprise in equivalence to *no nazdar!* resp. *no servus!*; *pomāli, pomāli!* 'not so fast!' < Moravian Czech or Slovak *pomaly* 'slow', etc. (cf. Newerkla 2007a, p. 281, 2007b, p. 41, 2013a, p. 256).

However, the 20th century also brought about a distinct change in the importance of the various Slavic minority groups in Vienna. Whereas the influence of Czech and Slovak inhabitants deteriorated, the importance of other groups increased (e.g. the Poles, Serbs, Croats and Bosnians, Turks, etc.). Linguistic consequences of this development are on the one hand the vanishing of several Czech and Slovak loanwords from the colloquial vocabulary of Viennese speakers, such as *Babutschen* 'fabric slippers' < *papuče*; *fix Laudon* 'blasted!' in

equivalence to *fix Laudon*; *geh' zum Tschert* 'go to hell!' < *jdi k čertu*; *Howno* 'shit' < *hovno*; *Klitsch* 'key', primarily in the sense 'skeleton key' < *klíč*; *Kudlička* 'simple penknife' < *kudlička*; *Mamlas* 'coward, idiot' < *mamlas*; *motz* 'much' < *moc*; *Naschi-Vaschi* 'a (forbidden) card play' < *naši – vaši* 'yours – ours'; *Nusch (N-/K-)* 'knife' < *nůž*; *Penise* 'money' < *peníze*; *Piwo* 'beer' < *pivo*; *platti/zaplatti* 'pay' < *platiti, zaplatiti*; *potschkai troschku* 'wait for a moment' < Moravian Czech and Slovak *počkaj trošku*; *(keinen) Rosomi haben* in the sense of 'have (no) wit' < *rozum* 'common sense'; *schezko jedno* 'no matter (who, what, when, where, why, how)' < *všecko jedno*; *Schwerak* 'comedian, rogue' < *čtverák*; *spatni* 'bad' < *špatný*; *Tamleschi* 'clumsy person' < *tam leží* '(s)he is lying there'; *Tanzowat* in denoting a dance club for Czech maids and soldiers < *tancovat* 'dance'; *Topanken* 'thick-soled ankle boots' < Slovak *topánky* 'shoes'; *Wetsch* 'button, small ball' < *veteš* 'junk, rubbish' in merging with *věc* 'thing'; *Wojak* 'soldier' < *voják*, etc. (cf. Newerkla 2009, p. 11, 2013a, p. 257). On the other hand, language contact with Bosnian, Croatian and Serbian has become the most important Slavic factor in influencing the colloquial language of many Viennese speakers of German, especially young ones, during the past years. Apart from these Southern Slavic languages, there are only two other languages with at least equally significant influence, i.e. English and Turkish.

2.2 Polycentric Convergence – The Example of Austria

Our second project part within the framework of the SFB "German in Austria. Variation – Contact – Perception" (F 60-G 23)[6] is called "German in the context of the other languages in the Habsburg state (19th century) and 2nd Austrian republic". The main goal of this part of the project is to provide a historically founded and multilingualism-based understanding of Austrian German's polycentricity. In this context, it seems appropriate to speak of polycentricity rather than pluricentricity (Clyne 1989, Ammon 1995, Schmidlin 2011, Auer 2013), since we are dealing with different historical factors in interaction with the centres of the Habsburg state that determined the status, functionality and structural heterogeneity of Austrian German. Beginning with the assumption that specific dimensions of – from this point of view – polycentric Austrian German are historically motivated, a central aim is to reconstruct the functional and metalinguistic dimensions of German in the multilingual Habsburg state and to relate them to the situation in the Second Austrian Republic.

---

6   See SFB "German in Austria. Variation – Contact – Perception": http://www.dioe.at (accessed 13/05/2019).

Investigating historical multilingualism in the Habsburg state can shed light on how, in the context of the other languages, German was used and valorised as an instrument of social interaction and as a reference point for cultural construction in East-Central Europe. Since communicative practices constitute a primary dimension of intercultural exchange, multilingualism represents a major signifier for non-national or multiple relations (e.g. Binder/Křivohlavá/Velek 2003, Evans 2004, Feichtinger/Cohen 2014, Goebl 1994, Judson 2006). Language as social practice provides access to ideologies and the ways people draw on, reproduce or create knowledge (cf. Heller 2001, pp. 214–215). This is true even more so for language ideologies and linguistic knowledge (e.g. Daskalov/Marinov 2013, Dorostkar 2014, Hentschel 1997). At the same time, historical multilingualism has had at least some impact on the linguistic structure of Austrian German (Newerkla 2013a, 2013b), but not much is known about aspects involving domain-specific communication or language-specific attitudes in the Habsburg state. So far, comparatively little language-centred historical research has been conducted on the interplay of officially imposed language regulations and unofficial multilingual practices in the domains of administration, the judiciary and education in the Habsburg monarchy, though such studies have been increasing since the 1990s (e.g. Burger 1995, Fellerer 2005, Newerkla 1999, Umberto/Rindler Schjerve/Metzeltin 1997). There were also ground-breaking research initiatives in this respect (Rindler Schjerve 2003) that explored how the struggle for power was reflected in attempts to control language use at different levels of discursive interaction and how, in the context of intricate and multiple language contact, language became a prominent site for interethnic controversies and conflicts.

Whereas the non-German-speaking nationalities of the Habsburg state attempted to redefine their status by demanding recognition of their languages and cultures, German-dominated state nationalism tried to re-establish its endangered hegemony by granting linguistic and cultural autonomy to the various ethnic groups. Hence, we hope that our investigations will yield new insights into the manner in which the different ethnic groups experienced the use of German – mediated through the multiple lingua-cultural practices – in their everyday lives. In addition, we will be able to understand how the diversity management from above and from below eventually shaped cultural encounters in East-Central Europe over time (Vetter 2003, Rindler Schjerve/Nelde 2003). In other words, we will try to identify the characteristics of the multilingual setting in which German was embedded at that time and which has most probably affected the language policies of the Second Austrian Republic as well as the

language behaviour of opinion leaders in the high-contact centres (most of all Vienna) – and thus German speakers in Austria – to the present day.

In this context, allow me to add a short aside in order to trigger even more thoughts on the question concerning the relationship between language, culture and society. If you belong to the deaf community in Austria, your perception of how languages are related in East-Central Europe usually differs decisively from our ordinary perception of language geography. Why is that? First, you most probably speak Austrian Sign Language (*Österreichische Gebärdensprache*, ÖGS), which is a fully-fledged natural language with complex structures and independent grammar as well as a sublexically significant sequential structure. "This means that, like spoken languages, sign languages have sub-lexical elements (phonology), morphology, semantics, syntax and pragmatics, and the lexicon consists of iconic and arbitrary signs." (cf. Krammer 2013, pp. 342–343). Second, for historical reasons Austrian Sign Language – together with Czech, Slovak and Hungarian Sign Languages – belongs to the language family of Austrian-Hungarian sign languages, which are part of the French Sign Language family. Also, the high degree of comprehensibility between the signed languages in Trieste (present-day Italy) and Austria is very probably due to their joint history of deaf schools within the Austro-Hungarian Empire (Dotter/Kellett Bidoli 2017). In contrast, although Austrian Sign language shares some aspects of its grammar with German and Swiss Sign Language, the vocabulary and thus the languages differ (Skant et al. 2002).

From the Austrian-Hungarian Sign Languages used in schools for the deaf in the Habsburg state, also the Slovenian and the Yugoslavian Sign Language (today Croatian Sign Language, Kosovar Sign Language, Serbian Sign Language) originated. And also the Russian Sign Language is said to have borrowed a lot of vocabulary from the Austrian-Hungarian sign languages due to the teachers in the first Russian schools for the deaf. In 1910, Russian Sign Language was also introduced in Bulgaria, where it has become a separate language (Bulgarian Sign Language) rather than a dialect of Russian Sign Language. However, whereas Wittmann (1991) classifies Bulgarian Sign Language as a descendent of Russian Sign, Bickford (2005) found that Bulgarian Sign formed a cluster with Slovak, Czech, Hungarian, Romanian, and Polish Sign. From this we can see that much research still has to be done from a historical point of view.

In Bulgaria, for example, the language of the classroom is different from that used by adults outside. Therefore it is not even clear, if Wittmann and Bickford looked at the same languages; nor, if one is derived from Russian Sign, if it is a dialect or if it creolised to form a new language. Not to speak of the above-mentioned Polish Sign Language that uses a one-handed manual alphabet based

on the alphabet from Old French Sign Language, whereas the language itself derives from German Sign Language (Farris 1994). Israeli Sign Language is also a descendant of German Sign Language, as it evolved from the sign language used by German Jewish teachers at a special school founded in 1873 by Marcus Reich. Several teachers from this school opened a school for deaf children in Jerusalem in 1932. Therefore, it still shows some resemblance to its German counterpart. But other sign languages or signing systems brought by immigrants also contributed to the emerging language, which started out as a pidgin. A local creole gradually emerged, which eventually became Israeli Sign Language. Today, this language is too removed from its origin to be considered a dialect of the German Sign Language.[7] Israeli Sign Language, however, is just the most commonly used sign language in the deaf community of Israel, where we can also find the Al-Sayyid Bedouin Sign Language[8] or a Hebrew manually coded language, and others.

As far as East-Central Europe is concerned, the question arises from time to time whether there was something like a Yiddish Sign Language. But as far as we know, there are no published descriptions or detailed attestations of its existence, although there may have been local varieties in pre-Holocaust Eastern Europe, especially in schools for the deaf. In *Glottolog 3.2.* (Hammarström/Forkel/ Haspelmath 2017), the entry on Yiddish Sign Language (Glottocode: yidd1241, ISO 639-3: yds) has been retired (effective from 2015-01-12). The justification for this step was that Yiddish Sign Language was "non-existent". As Bernard Spolsky (2014) in his entry to the Jewish Language Research Website aptly remarks, experts in Sign Language have not heard of Yiddish Sign Language, neither Wendy Sandler, nor Nancy Brunlehrman, nor Bram Weiser nor Adele Kronick Shuart. But he continues, "There was a school in Cracow, the Yiddishe Toib Shtim Shule, where the pupils probably used a Sign Language amongst themselves (even though the school officially used spoken Yiddish). Mark Zaurov, a Deaf historian studying the experiences of the Jewish Deaf in the Holocaust, found mentions of several Deaf Jewish schools where many children spoke Yiddish; they may have had a local sign language." (cf. Spolsky 2014, para. 4) Looking back, it is obviously quite difficult to come up with relevant and accurate data in this respect. Nevertheless there "may have been distinctive sign languages used by Deaf communities in Eastern Europe before the war. But a distinct unified Yiddish Sign Language is unattested and unlikely." (cf. Spolsky 2014, para. 4)

---

7   For the detailed story of Israeli Sign Language see Meir/Sandler (2008).
8   For more information on this language see Meir/Sandler/Padden/Aronoff (2010).

3 Conclusion

Our aim was to provide the reader with some interesting glimpses into the problems that make analysing linguistic areas in East-Central Europe, as the result of pluridimensional, polycentric convergence phenomena, a real challenge. Much research has to be done on that matter, much has already been achieved, but much is still ahead of us on the way to a better understanding of linguistic areas as such and especially in East-Central Europe. From the standpoint of contact linguistics and historical sociolinguistics, we should always bear in mind that languages do not converge by themselves, but that it is the agency or unconscious behaviour of speakers that has this effect (cf. Gijn/Muysken 2016, s. p.). In this context, we still must investigate further into the contribution of specific groups of people to the emergence of certain areas. In East-Central Europe, for example, large numbers of Jews identified with an ideal vision of German *Bildung* and enlightenment. "[…] the concept of *Bildung* became for many Jews »synonymous with their Jewishness.« It would be a fundamental instrument of cultural integration into German middle-class society in Austria. The German language and culture also provided a gateway to economic advancement and rising social status in the Austro-Hungarian Monarchy. This was a crucial factor that influenced Jews as far apart as Bohemia, Hungary, Bukovina and the Adriatic port of Trieste." (cf. Wistrich 2007, p. 58) Thereby, they helped to establish German as the *lingua franca* of the polyglot monarchy, and, together with the state officials and the army, laid the foundations for the Habsburg empire to become a linguistic area in East-Central Europe with certain characteristics. In this context, the role of Yiddish as a means to bridge the gap between Austrian German and the various languages of the monarchy, the Slavic languages in particular, has not yet drawn proper scientific attention. Uncovering systematically the hidden multilingualism of that time is still an important desideratum of research in the field, although we are often confronted with the lack of sufficiently meaningful data.

Furthermore, as we have shown in the second part of our paper, in this context it seems more appropriate to speak of polycentric rather than pluricentric convergence. Following Clyne (1989), polycentricity is not entirely the same as pluricentricity, because "the latter term emphasizes plurality of varieties within a language, i.e. plurality of relatively stable self-contained linguistic systems that together make up a language. Polycentricity emphasizes the functional inequality between such varieties and the simultaneous links to the various centering powers language practices are simultaneously subject to. Whereas a pluricentric language is the sum of its varieties, a polycentric language is a dynamic, socially

ordered system of resources and norms that are strongly or weakly associated with one or more centers." (cf. Li/Juffermans 2012, 77). For instance, as we could see from the example of the Austrian-Hungarian sign languages, their different evolution and history, these kinds of signed languages are not so tied to the spoken languages of a region, but rather to a place or a social stratum. This fact makes the concept of linguistic areas appear even more vivid and dependent on social interaction, e.g. at schools or in other language domains, rather than on the specific characteristics of the languages in contact. Again, as said above, languages do not converge by themselves, it is the behaviour of the speakers that brings about these pluridimensional, polycentric convergence phenomena leading to specific linguistic areas. While this phenomena is not exclusive to East-Central Europe, it is certainly applicable to it.

References

Ammon, Ulrich: *Die deutsche Sprache in Deutschland, Österreich und der Schweiz. Das Problem der nationalen Varietäten*. De Gruyter: Berlin/New York. 1995.

Auer, Peter: Enregistering pluricentric German. In: Augusto Soares da Silva (ed.): *Pluricentricity. Language Variation and Sociocognitive Dimensions*. De Gruyter: Berlin/Boston. 2013, pp. 19–48.

Bickford, J. Albert: *The Signed Languages of Eastern Europe* (= SIL Electronic Survey Reports 2005, 26). SIL International: Tucson, AZ/University of North Dakota: Grand Forks, ND. 2005.

Binder, Harald/Barbora Křivohlavá/Luboš Velek: *Místo národních jazyků ve výchově, školství a vědě v habsburské monarchii 1867–1918/Position of National Languages in Education, Educational System and Science of the Habsburg monarchy, 1867–1918*. Výzkumné centrum pro dějiny vědy: Praha. 2003.

Blaha, Ondřej: *Jazyky střední Evropy*. Univerzita Palackého: Olomouc. 2015.

Burger, Hannelore: *Sprachenrecht und Sprachengerechtigkeit im österreichischen Unterrichtswesen* 1867 –1918 (= Studien zur Geschichte der österreichischen Monarchie 26). Verlag der österreichischen Akademie der Wissenschaften: Wien. 1995.

Clyne, Michael: Pluricentricity: National Variety. In: Ammon, Ulrich (ed.): *Status and Function of Languages and Language Varieties*. De Gruyter: Berlin/New York. 1989, pp. 357–371.

Daskalov, Roumen/Marinov, Tchavdar (eds.): *Entangled Histories of the Balkans. Vol. 1: National Ideologies and Language Policies* (= Balkan Studies Library 9). Brill Academic Publishers: Leiden. 2013.

Dorostkar, Niku: *(Mehr-)Sprachigkeit und Lingualismus. Die diskursive Konstruktion von Sprache im Kontext nationaler und supranationaler Sprachenpolitik am Beispiel Österreichs* (= Kommunikation im Fokus – Arbeiten zur Angewandten Linguistik 3). V&R Unipress: Göttingen. 2014.

Dotter, Franz/Bidoli, Cynthia J. Kellett: The Historical Relationship between Triestine Sign Language and Austrian Sign Language. In: *Sign Language Studies* 17(2), 2017, pp. 193–221.

Ernst, Peter: Die Tschechen in Wien und ihr Einfluss auf das Wienerische: eine kritische Bestandsaufnahme. In: Nekula, Marek/Bauer, Verena/Greule Albrecht (eds.): *Deutsch in multilingualen Stadtzentren Mittel- und Osteuropas*. Praesens Verlag: Wien. 2008, pp. 99–107.

Evans, Robert J. W: Language and State Building: The Case of the Habsburg Monarchy. In: *Austrian History Yearbook* 35, 2004, pp. 1–24.

Farris, Michael A: Sign language research and Polish sign language. In: *Lingua Posnaniensis* 36, 1994, pp. 13–36.

Feichtinger, Johannes/Cohen, Gary B. (eds.): *Understanding Multiculturalism. The Habsburg Central European Experience*. Berghahn Books: New York/Oxford. 2014.

Fellerer, Jan: *Mehrsprachigkeit im galizischen Verwaltungswesen (1772 –1914): eine historisch-soziolinguistische Studie zum Polnischen und Ruthenischen (Ukrainischen)* (= Bausteine zur Slavischen Philologie und Kulturgeschichte: Reihe A, Slavistische Forschungen, N. F. 46). Böhlau-Verlag: Köln/Wien. 2005.

Gijn, Rik van/Muysken, Pieter: Linguistic Areas. In: *Oxford Bibliographies Online*. 2016. Retrieved at 13.05.2019, from http://dx.doi.org/10.1093/obo/9780199772810-0133.

Goebl, Hans: Geschichte lernen und aus Geschichte lernen: Die altösterreichische Sprachenvielfalt und -politik als Modellfall für ein Europa von heute und morgen. In: Helfrich, Uta/Riehl, Claudia Maria (eds.): *Mehrsprachigkeit in Europa – Hindernis oder Chance?* Egert: Wilhelmsfeld. 1994, pp. 55–82.

Goebl, Hans/Nelde, Peter H./Starý, Zdeněk/Wölck, Wolfgang (eds.): *Kontaktlinguistik. Contact Linguistics. Linguistique de contact. Ein internationales Handbuch zeitgenössischer Forschung. An International Handbook of Contemporary Research. Manuel international des recherches contemporaines* (= Handbücher zur Sprach- und Kommunikationswissenschaft 12). 2 vol. De Gruyter: Berlin/New York. 1996–1997

Gold, David L.: Zinfandel: An American English Grape and Wine Name of Immediate Hungarian, Moravian Czech, and/or Slovak Origin (On

How the Origin of a Significans Need Not Be Parallel to the Origin of the Corresponding Significandum). In: Gold David L. (ed.): *Studies in Etymology and Etiology. With Emphasis on Germanic, Jewish, Romance, and Slavic Languages*. Selected and edited, with a foreword, by Félix Rodríguez González and Antonio Lillo Buades. Publicacions Universitat Alacant: Alicante. 2009, pp. 669–708.

Grüner, Sigmar/Sedlaczek, Robert: *Lexikon der Sprachirrtümer Österreichs*. Anaconda: Wien. 2003.

Hammarström, Harald/Forkel Robert/Haspelmath, Martin (eds.): *Glottolog 3.0*. 2017. Retrieved at 13.05.2019, from http://glottolog.org/.

Heller, Monica: Undoing the macro/micro dichotomy: Ideology and categorisation in a linguistic minority school. In: Coupland, Nikolas/Sarangi, Srikant/Candlin, Christopher N. (eds.): *Sociolinguistics and Social Theory*. Routledge: London. 2001, pp. 212–234.

Hentschel, Gerd (ed.): *Über Muttersprachen und Vaterländer. Zur Entwicklung von Standardsprachen und Nationen in Europa*. Frankfurt am Main: Peter Lang 1997.

Hymes, Dell: *Foundations in Sociolinguistics. An Ethnographic Approach*. University of Pennsylvania Press: London. 1974.

Jakobson, Roman : Sur la théorie des affinités phonologiques des langues. In: *Actes du quatrième congrès international de linguistes tenu à Copenhague du 27 août au 1er septembre 1936*. Munksgaard : Copenhague. 1938, pp. 48–59.

Januška, Jiří: *Porovnávání středoevropských jazyků: za horizont strukturních rysů a lexikálních přejímek* [Comparing Central European languages: beyond structual features and loanwords]. Ph.D. dissertation, Ústav obecné lingvistiky, Filozofická fakulta, Univerzita Karlova. 2017.

Judson, Pieter M.: *Guardians of the Nation. Activists on the Language Frontiers of Imperial Austria*. Harvard University Press: Cambridge, MA. 2006.

Krammer, Klaudia: The benefits of sign language for deaf children with and without cochlear implant(s). In: *European Scientific Journal* 4, 2013, pp. 341–349.

Li, Jinling/Juffermans, Kasper: Chinese complementary schooling in the Netherlands: experiences and identities of final-year students. In: Grande, Francesco/Ruiter, Jan Jaap de/Spotti, Massimiliano (eds.): *Mother tongue and intercultural valorization: Europe and its migrant youth*. FrancoAngeli: Milano. 2012, pp. 61–80.

Masařík, Zdeněk: Zum tschechisch-deutschen/österreichischen Sprachkontakt (dargestellt am Beispiel Südmähren – Niederösterreich. In: Bauer,

Werner/Scheuringer, Hermann (eds.): *Beharrsamkeit und Wandel. Festschrift für Herbert Tatzreiter zum 60. Geburtstag*. Ed. Praesens: Wien. 1998, pp. 133–141.

Meir, Irit/Sandler, Wendy: *A Language in Space. The Story of Israeli Sign Language*. Psychology Press: New York/London. 2008.

Meir, Irit/Sandler, Wendy/Padden, Carol/Aronoff, Carol/Aronoff, Mark: Emerging sign languages. In: Marschark, Marc/Spencer, Patricia Elizabeth (eds.): *Oxford Handbook of Deaf Studies, Language, and Education*. Volume 2. Oxford University Press: Oxford. 2010, pp. 267–280.

Newerkla, Stefan Michael: *Intendierte und tatsächliche Sprachwirklichkeit in Böhmen. Diglossie im Schulwesen der böhmischen Kronländer 1740 –1918* (= Dissertationen der Universität Wien 61). WUV: Wien. 1999.

Newerkla, Stefan Michael: Language Affinity in Central Europe – Some Thoughts on the Interrelations of German, Czech, Slovak and Magyar. In: *Opera slavica. Slavistické rozhledy* 10(4), 2000, pp. 1–16.

Newerkla, Stefan Michael: Areály jazykového kontaktu ve střední Evropě a německočeský mikroareál ve východním Rakousku. [Language contact areas in Central Europe and the German-Czech micro-area in Eastern Austria]. In: *Slovo a slovesnost* 68, 2007a, pp. 271–286.

Newerkla, Stefan Michael: Kontaktareale in Mitteleuropa. [Areały kontaktów językowych w Europie Środkowej]. In: Kątny, Andrzej (ed.): *Słowiánsko-niesłowiańske kontakty językowe. Slawisch-nichtslawische Sprachkontakte. Słowiańsko-niesłowiańske kontakty językowe w perspektywie dia- i synchronicznej. Slawisch-nichtslawische Sprachkontakte in dia- und sychronischer Sicht*. Wydawnictwo Wszechnicy Mazurskiej: Olecko. 2007b, pp. 29–48.

Newerkla, Stefan Michael: Teebutter, Teewurst, Thea und der Tee. In: *ÖGL – Österreich in Geschichte und Literatur (mit Geographie)* 52(4–5a), 2008, pp. 240–252.

Newerkla, Stefan Michael: Bohemismen (und Slowakismen) in Wien (und Österreich). In: *Zeitschrift für Sprache und Schreibung* 3, 2009, pp. 8–12.

Newerkla, Stefan Michael: *Sprachkontakte Deutsch – Tschechisch – Slowakisch. Wörterbuch der deutschen Lehnwörter im Tschechischen und Slowakischen: historische Entwicklung, Beleglage, bisherige und neue Deutungen*. 2nd edn. (= Schriften über Sprachen und Texte 7). Peter Lang: Frankfurt am Main. 2011. Retrieved at 13.05.2019, from http://dx.doi.org/10.3726/978-3-653-03121-8.

Newerkla, Stefan Michael: Čeština ve Vídni na konci 18. století. [Czech in Vienna at the end of the 18th century]. In: Čmejrková, Světla/Hoffmannová,

Jana/Klímová, Jana (eds.): *Čeština v pohledu synchronním a diachronním*. Karolinum: Praha. 2012a, pp. 241–246.

Newerkla, Stefan Michael: Kontinuität und Wandel in der Minderheitensituation – Tschechisch in Wien bis 1775. In: Woldt, Claudia (ed.): *Tschechisch bis 1775 – historische Kontinuität oder Geschichte mit Sollbruchstellen?* (= Specimina Philologiae Slavicae 169). Peter Lang: München/Berlin/Washington D.C. 2012b, pp. 67–80.

Newerkla, Stefan Michael: Pischinger, Zierfandler, Kaisersemmel und Zuckerkandl: Altösterreich als kulinarisches Sprachareal. In: Bergmann, Hubert/Unterguggenberger, Regina Maria (eds.): *Linguistica culinaria. Festgabe für Heinz-Dieter Pohl zum 70. Geburtstag*. Praesens Verlag: Wien. 2012c, pp. 309–323.

Newerkla, Stefan Michael: Linguistic Consequences of Slavic Migration to Vienna in the 19th and 20th Centuries. In: Moser, Michael/Polinsky, Maria (eds.): *Slavic Languages in Migration* (= Slavische Sprachgeschichte 6). LIT: Berlin/Wien. 2013a, pp. 247–260.

Newerkla, Stefan Michael: Vom Nachwirken einer Minderheit: Zuwanderer aus den böhmischen Ländern in Wien. In: *Wiener Slavistisches Jahrbuch. Neue Folge – Vienna Slavic Yearbook. New Series* 1, 2013b, pp. 1–19.

Pohl, Heinz Dieter: Zum österreichischen Deutsch im Lichte der Sprachkontaktforschung. In: *Klagenfurter Beiträge zur Sprachwissenschaft* 25, 1999, pp. 93–115.

Pohl, Heinz Dieter: *Die österreichische Küchensprache*. Praesens Verlag: Wien. 2007.

Rindler Schjerve, Rosita (ed.): *Diglossia and Power: Language Policies and Practice in the 19th Century Habsburg Empire* (= Language, Power and Social Process 9). De Gruyter: Berlin/New York. 2003.

Rindler Schjerve/Nelde, Rosita/Nelde, Peter H. (eds.): *Der Beitrag Österreichs zu einer europäischen Kultur der Differenz. Sprachliche Minderheiten und Migration unter die Lupe genommen*. Asgard-Verlag: St. Augustin. 2003.

Schmidlin, Regula: *Die Vielfalt des Deutschen: Standard und Variation. Gebrauch, Einschätzung und Kodifizierung einer plurizentrischen Sprache*. De Gruyter: Berlin/New York. 2011.

Schuster, Mauriz/Schikola Hans: *Das alte Wienerisch. Ein kulturgeschichtliches Wörterbuch*. Deuticke: Wien. 1996.

Sedlaczek, Robert: *Kleines Handbuch der bedrohten Wörter Österreichs.* Unter Mitarbeit von Melita Sedlaczek. Wirtschaftsverlag Ueberreuter: Wien. 2007.

Sedlaczek, Robert: *Wörterbuch des Wienerischen*. Haymon: Wien. 2011.

Skála, Emil: Tschechisch-deutsche Sprachkontakte. In: *Přednášky z XLI. běhu Letní školy slovanských studií*. Univerziteta Karlova: Praha. 1998, pp. 213–227.

Skant, Andrea et al.: *Grammatik der Österreichischen Gebärdensprache* (= Veröffentlichungen des Zentrums für Gebärdensprache und Hörbehindertenkommunikation, Band 4). Forschungszentrum für Gebärdensprache und Hörgeschädigtenkommunikation: Klagenfurt. 2002.

Spolsky, Bernard: Yiddish Sign Language. In: Benor, Sarah Bunin (ed.): *Jewish Language Research Website*, 2014. Retrieved at 24.01.2014, from http://www.jewish-languages.org/yiddish-sign-language.html.

Steinhauser, Walter: *Slawisches im Wienerischen* (= Schriftenreihe des Vereines „Muttersprache" Wien 7). Notring der Wissenschaftlichen: Wien. 1978.

Thomas, George: The Role of German Loanwords in the Slavic Languages of the Former Habsburg Empire. In: *Canadian Slavonic Papers/Revue canadienne des slavistes* 39(3–4), 1997, pp. 333–359.

Thomason, Sarah Grey: Linguistic areas and language history. In: Gilbers, Dicky G./Nerbonne, John/Schaeken, Jos (eds.): *Languages in Contact* (= Studies in Slavic and General Linguistics, vol. 28). Brill Academic PUB: Amsterdam/Atlanta, GA. 2000, pp. 311–327.

Umberto, Rinaldi/Rindler-Schjerve, Rosita/Metzeltin, Michael (eds.): *Lingua e politica. La politica linguistica della duplice monarchia e la sua attualità. Sprache und Politik. Die Sprachpolitik der Donaumonarchie und ihre Aktualität. Atti del simposio Istituto Italiano di Cultura, 31. 5. 1996.* Red.: Gualtiero Boaglio. Instituto Italiano di Cultura: Vienna. 1997.

Vetter, Eva: Hegemonic discourse in the Habsburg empire: the case of education. A critical discourse analysis of two mid 19th century government documents. In: Rindler-Schjerve, Rosita (ed.): *Diglossia and Power: Language Policies and Practice in the 19th Century Habsburg Empire* (= Language, Power and Social Process 9). De Gruyter: Berlin/New York. 2003, pp. 271–307.

Wehle, Peter: *Sprechen Sie Wienerisch? Von Adaxl bis Zwutschkerl*. Wirtschaftsverlag Ueberreuter: Wien. 1980.

Wehle, Peter: *Sprechen Sie ausländisch? Von Amor bis Zores*. 2nd edn. Wirtschaftsverlag Ueberreuter: Wien. 1996.

Wehle, Peter: *Die Wiener Gaunersprache. Von Auszuzln bis Zimmerwanzen*. 2nd edn. Wirtschaftsverlag Ueberreuter: Wien. 1997.

VoGrG: *Bundesgesetz vom 7. Juli 1976 über die Rechtsstellung von Volksgruppen in Österreich (*Volksgruppengesetz). StF: BGBl. Nr. 396/1976 (NR: GP XIV

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

---

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

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'.

(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 *stuždь* or *štuždь* '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 *ǫgorъ. 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 *brašьno* '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 *$h_1oh_1\acute{k}u$-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. *rǎka* '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. *jabǎlka* '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-eh₂* or **als-eh₂* (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 **gʷeh₂-* (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 **ar̯dla* > 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_, *_ǫgorʾ_). 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

---

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.

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.

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: Probleme 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.

Thomas Stolz & Nataliya Levkovych

On Different Ways of Belonging in Europe

Abstract: In this study we investigate the hitherto largely neglected issue of the areal distribution of BELONG-constructions in Europe. Several isoglosses are identified on the basis of the etymology of the lexical verb that is used to express the notion of BELONG. It is shown that these isoglosses divide Europe in different subareas which meet in East-Central Europe. We also discuss evidence for the diffusion of the typically Germanic HEARING-isogloss into the territory of neighboring Baltic, Slavic, and Uralic languages. Furthermore, we show that substantival possessive constructions are attested (as expected) in all our sample languages. It is checked whether there are any correlations of BELONG-constructions and genetic affiliation or the presence/absence of lexical HAVE-verbs in a given language. The data are indicative of a new area of research which holds many interesting new insights in store for students of possession.

Keywords: areal linguistics, possessive constructions, language contact, European isoglosses, typology

1 Introduction

This paper forms part of a research program dedicated to the typological assessment of hitherto largely understudied predicative possessive notions, excluding the relatively well-established category of HAVE-possession (= H_{poss}) (Stolz/Levkovych 2017). The focus of this study is on BELONG-possession constructions (= B_{poss}) which obtain in the languages of Europe. The principal aim of the paper is to identify the areas in which the different kinds of B_{poss}-constructions are attested. Furthermore, it demonstrates that some of the B_{poss}-constructions have been expanding in the sense that they have been copied either materially or as calques as languages have come into contact. In conformity with the general orientation of this edited volume, special attention is paid to scenarios, which involve languages from the East-Central Europe.

This paper reports on the earliest stages of our only recently initiated project and, thus, many open questions remain to be tackled in the future. Except for English, French, German, and Ukrainian examples, for which we claim sufficient foreign language and native language competence, respectively, the data on which our line of argumentation rests are taken exclusively from the extant descriptive-linguistic literature, i.e., mostly grammars and dictionaries of

sixty-eight European languages.[1] Therefore, standard varieties form the backbone of our empirical documentation. Nonstandard varieties are only unsystematically addressed, if at all. The methodology of our choice is strictly qualitative.

Since the sources on which we rely very often do not explicitly make statements as to the (non-)existence of B_{poss}-constructions, we are cautious not to jump to conclusions on still relatively shaky empirical foundations. This means that for a sizable number of languages, we refrain from classifying them once and for all. The geographic distribution of certain B_{poss}-construction types is represented on a linguistic map of Europe in Appendix I. The languages which are represented on this map are listed separately with additional information in Appendix II. For the purpose of this study, we consider Europe in the terms of König/Haspelmath (1999, pp. 112–114). In this sense the Cis-Uralian part of Kazakhstan, the entire (Trans-)Caucasian region, Anatolia, Cyprus, Malta, and Iceland are counted in.

To achieve the goals identified in the initial paragraph we first provide the necessary background information as to the current debate within the framework of linguistic research on possession (= Section 2). Section 3 deals with those European languages for which the existence of genuine B_{poss}-constructions is either explicitly denied or doubtful. Section 4 is dedicated to the presentation and discussion of B_{poss}-constructions as attested in Europe. In the same section, the issue of the transferal of patterns of B_{poss}-constructions in situations of language contact is addressed. In Section five we present our papers' conclusions.

2 Background

Linguistic research on possession looks back on a rich and venerable tradition starting with the seminal work of Seiler's (1973), followed by e.g., Heine's (1997) influential monograph, and culminating in Stassen's (2009) monumental typological overview of predicative possession in the languages of the world, to mention only three of the most prominent representatives of this research program. The two major subdivisions of possession are often inquired into separately from each other in dedicated scholarly work, as is the case with Haspelmath (2017) on adnominal and Mazzitelli (2015) on predicative possession. Many of the studies within the domain of predicative possession published over the last forty years focus on so-called H_{poss}-constructions.

H_{poss}-constructions describe possessive situations, which involve a possessor as the participant about whom it is predicated that s/he has X with X being the

---

1   To save space, the sources of single word examples in the main body of the text are not given.

second participant, viz. the possessee. The predicate nucleus functions as relator of the two participants and, at the same time, specifies the kind of possessive relation that exists between the possessor and possessee. A typical H_{poss}-constructions from English is given in (1) to identify the properties which are commonly held to be characteristic of this kind of predicative possessive constructions.

(1)   English: H_{poss}-construction
       [The child]$_{possessor}$ [has]$_{relator}$ [a toy]$_{possessee}$

Cross-linguistically, H_{poss}-constructions come in a limited range of shapes for which cognitively anchored conceptual schemas have been postulated (Heine 1997, p. 47). Moreover, the geographic distribution of the different types of constructions that realize these schemas (or a modified set thereof) yields patterns, which are suggestive of a certain degree of areality (Stassen 2005, pp. 476–477). According to the maps that capture the global situation, Europe is special insofar as it hosts numerous languages, which boast a proper, i.e., lexical H_{poss}-verb (LEXH_{poss}-verb)—a category, which is clearly a minority option outside of the European continent. However, in Europe there are numerous competitors of LEXH_{poss}-verbs too, namely constructions that reflect spatial, genitival, comitative, or other schemas. In Mazzitelli (2017) and Levkovych/Mazzitelli/Stolz (accepted) areal aspects of the distribution of the different HAVE-constructions over the languages of the Circum-Baltic area and across the entire European continent are charted, respectively. The bulk of the languages which attest to LEXH_{poss}-verbs occupy the vast territory in most of the western half of the continent, whereas languages which employ different means for the expression of H_{poss} are situated predominantly in the eastern part of Europe, although there are also the Celtic languages on the extreme western fringes whose system of predicative possession excludes a proper H_{poss}-verb. Our knowledge of the geo-linguistics of H_{poss} in Europe is therefore already sufficiently detailed.

In contrast to the impressively rich literature on H_{poss}, there is as yet only a relatively insignificant output of linguistic work addressing B_{poss} in-depth. At times, B_{poss} is explicitly excluded from work on predicative possession because the notion is believed to be lacking in prototypicality (Mazzitelli 2017, p. 8). Stassen (2009, p. 11) complains that "[t]he notion of 'belonging' is of course pre-theoretical and vague." This problematic state of affairs can only be remedied if our empirical database of what cross-linguistically counts as an instance of B_{poss} considerably gains in size. The almost complete absence of dedicated studies on this issue is quite surprising, because H_{poss} and B_{poss} are usually depicted as fraternal twin notions (Heine 1997, pp. 29–30), in a manner of speaking. Their distinction

by formal means is explicitly believed to be universal not only by Heine (1997, pp. 32–33). According to a widely shared opinion, B_{poss}-constructions picture possessive situations differently from H_{poss}-constructions in the sense that, in the case of B_{poss}, the predication is about the possessee, which is ascribed to the possessor by the predicate nucleus. This means that the orientation of the predication is the reverse of that of H_{poss}-constructions—an aspect of possession with which several authors have taken issue (e.g., Seiler 1983, p. 61; Stolz et al. 2008, pp. 20–24). Aikhenvald (2013, p. 29) assumes that the distinction of H_{poss} vs. B_{poss} is basically of a pragmatic nature whereas Mazzitelli (2015, p. 38) argues that "[t]he distribution of pragmatic roles also fails to offer a reliable universal criterion".

To cut a long discussion short, we transform the above template of the H_{poss}-construction in (1) into that of a typical English B_{poss}-construction in (2). As a matter of fact, the predicator is complex as the (transitive) prepositional verb consists of the lexical (and originally intransitive) verb *belong* plus the allative preposition *to* which otherwise links indirect objects or spatial adjuncts to the predicate nucleus.

(2)   English: belong-constructions
        [The toy]$_{possessee}$ [belongs to]$_{relator}$ [the child]$_{possessor}$

In purely structural terms, English H_{poss}-constructions and B_{poss}-constructions are clearly different from each other. However, this does not mean that this is generally the case in human languages. Khizanishvili (2006, pp. 16–17) for instance, discusses the distinction in Georgian (which is one of our European sample languages) and concludes that morpho-syntactically the two constructions cannot be told apart easily. To illustrate this problem we reproduce Khizanishvili's examples in (3)-(4).

(3)   Georgian: H_{poss}-construction (Khizanishvili 2006, p. 16)
        me   ċign-i   m-akv-s
        1sg.dat   book-nom   1sg.obj-have-3sg.sbj
        'I have a book.'

(4)   Georgian: B_{poss}-construction (Khizanishvili 2006, p. 17)
        me   ċign-i   m-e-ḳutvni-s
        1sg.dat   book-nom   1sg.obj-chv-belong-3sg.sbj
        'I have a book.'

In both (3) and (4), the possessor is represented by the pronoun of the first person singular in the dative whereas the possessee comes in the shape of a

lexical noun in the nominative. The verbal agreement morphology accordingly reflects the subject status of the possessee and the object status of the possessor. As to the object function of the possessor, however, the two sample sentences differ. In (3), the possessor is the direct object, whereas in (4) the same pronoun *me* 'me' functions as indirect object, which can only be seen from the use of the so-called character vowel *-e-* on the right of the object prefix of the finite verb (Fähnrich 1986, p. 73). Thus, the two constructions fail to be absolutely identical. Nevertheless, the possessee-NP always fulfills the function of subject, whereas the possessor is excluded from this fundamental relation. Most probably, both Georgian constructions would be considered instances of the Goal Schema by Heine (1997, p. 59–61) whereas Stassen (2009, p. 292–294) classifies the Georgian H_{poss}-construction as an instance of his Locational Possessive. Thus, there is disagreement as to the conceptual interpretation of the constructions under review, no matter whether they count as examples of H_{poss} or B_{poss}.

On top of that, B_{poss}-constructions do not constitute a homogeneous class. In fact, there are two major types, namely substantival possessive (= S_{poss}) constructions (Ultan 1978) and lexical B_{poss}-constructions (= LEXB_{poss}) (Stolz/ Levkovych accepted). In the main part of this study, we search for evidence of the latter kind but cannot help mentioning S_{poss}-constructions repeatedly. We sketch the characteristic properties of English S_{poss}-constructions in (5)-(6).

(5)　English: S_{poss}/lexical possessor (genitive)
　　　[The toy]$_{possessee}$ [is]$_{copula}$ [the child]$_{possessor}$-[′s]$_{relator}$

(6)　English: S_{poss}/pronominal possessor (possessive pronoun)
　　　[The toy]$_{possessee}$ [is]$_{copula}$ [her]$_{possessor}$-[s]$_{relator}$

First of all, S_{poss}-constructions are claimed to be universal by Clark (1978, p. 90)—no such claim has been made with reference to LEXB_{poss}-constructions (and for good reason). The difference between the construction in (2) and those in (5)-(6) hinges on the choice of predicator or relator. In (2), the predicate nucleus is a lexical verb, whereas in (5)-(6) we find a copula in this position.

Examples like (2) from English motivate the terminological choice we make in the linguistics of possession. As Aikhenvald (2013, p. 29) rightfully criticizes, Eurocentric labels for possessive categories are treacherous because they might invoke associations which are not in line with the structural facts of (not only extra-European) languages. In Swahili for instance, there is no LEXB_{poss}-verb, in the first place. The examples (7)-(9) illustrate how the concepts of H_{poss} and B_{poss} are expressed conventionally in this Bantu language.

(7)   Swahili: H$_{poss}$ (Kwon 1995, p. 177)
      M-toto   a-na   kitabu
      CL1-child   3SG-with   (CL7)book
      'The child has a book.'

(8)   Swahili: S$_{poss}$/lexical possessor (Kwon 1995, p. 177)
      Kitabu   ni   ch-a   m-toto
      (CL7)book   COP   CL7-CONN   CL1-child
      'The book belongs to the child.'

(9)   Swahili: S$_{poss}$/pronominal possessor (Kwon 1995, p. 180)
      Nyumba   hi-i   (ni)   y-ako
      (CL9)house   DEM-CL9   (COP)   CL9-POSS.2SG
      'This house is yours.'

What we have in Swahili are constructions, which lack any proper verbal predicator. There is neither a transitive H$_{poss}$-verb nor is there a LEXB$_{poss}$-verb. The recent literature on this subject matter contains ever more statements which deny the existence of LEXB$_{poss}$-verbs in certain languages, such as e.g., Nêlêmwa (Bril 2013, p. 85). Accordingly, Dixon (2012, p. 302) concludes that having a LEXB$_{poss}$-verb is the privilege of "a small minority" of the world's languages. In a way, the term is thus a misnomer because it presupposes that human languages employ patterns, which are particular to European languages such as English. If it is true that the category of LEXB$_{poss}$-verbs fails to be properly universal, which conclusions can be drawn about predicative possession in those languages, which nevertheless give evidence of verbs of this kind? To settle this issue once and for all, we are in dire need of sufficiently large empirical foundations—meaning: we still have to take stock of B$_{poss}$-constructions for as many languages as possible in order to determine in what way and to what extent the presence of B$_{poss}$-constructions in a given language correlates with other properties of the same language. In this study, we take a relatively small step in this direction by way of reviewing how LEXB$_{poss}$-verbs fare on the European continent, where they are believed to thrive.

3 On the Absence of a *LEXB*$_{poss}$-Verb

We open the empirical overview by way of looking at the data from languages, which lack a LEXB$_{poss}$-verb. LEXB$_{poss}$-verbs are not attested in 21 of the 68

languages taken into account in this study, i.e., slightly less than 31 % of our sample is devoid of a verb of this kind. Since in the traditional canon of European grammar writing there is no chapter reserved for this class of verbs, we cannot be sure that the grammarian's silence on this issue is tantamount to stating the absence of a $\text{LEXB}_{\text{poss}}$-verb from a given language. In case of doubt, we therefore classify these languages as potentially lacking a $\text{LEXB}_{\text{poss}}$-verb. It is an urgent task for follow-up studies to clarify these cases.

We divide this section in two parts according to the genealogy of the object languages. For a start, we look at the absence of $\text{LEXB}_{\text{poss}}$-verbs in non-Indo-European languages of Europe in Section 3.1. The same phenomenon as attested in Indo-European languages of Europe is discussed in Section 3.2. To save space we restrict the presentation of sentential examples to a minimum, i.e., a given language is chosen as representative of a group of languages with similar properties.

3.1 Non-Indo-European Languages without $\text{LEXB}_{\text{poss}}$-Verbs

The isolate Basque not only has a proper H_{poss}-verb (*edun ~ eduki ~ ukan*) but also makes use of constructions of S_{poss}-construction. In (10), we exemplify the H_{poss}-construction with a sentence from the standardized variety of Basque (Batúa) whereas for the S_{poss}-verb in (11) we provide an example from the Labourdin variety of Basque.

(10)   Basque: H_{poss} (Hualde/Ortiz de Urbina 2003, p. 221)
       Diru-a   dut
       money-DEF   have:1SG.ERG
       'I have money.'

(11)   Basque: S_{poss} (Lafitte 1998, p. 60)
       Liburu   hori   Piarres-en-a   da
       book   DEM.PROX   Pierre-GEN-DEF   be.3SG.ABS
       'This book is Pierre's.'

The Basque examples demonstrate two things, namely that $\text{LEXB}_{\text{poss}}$-verbs

(a) may be absent from a modern European language
(b) even in the presence of a proper $\text{LEXH}_{\text{poss}}$-verb in the system of predicative possession of the same language.

One might want to argue that Basque frequently structurally differs from the bulk of the languages of Europe—especially those of the Indo-European language

family. Accordingly, the absence of a lexB$_{poss}$-verb might be added to the list of Basque idiosyncrasies in the European context.

However, a similar situation can be found in numerous non-Indo-European languages spoken in the eastern regions of the continent. The absence of a lexB$_{poss}$-verb is pervasive in the Turkic languages of Europe. The Chuvash examples in (12)-(13) reflect a general Turkic pattern.

(12)   Chuvash: H$_{poss}$ (Landmann 2014a, p. 29)
        Man(-ăn)   kĕneke-m   pur.
        1sg(-gen)   book-por.1sg   exi
        'I have a book.'

(13)   Chuvash: S$_{poss}$ (Landmann 2014a, p. 13)
        Ku   çurt   kürš-ĕn.
        dem.prox   house   neighbor-gen
        'This house is the neighbor's.'

As in the above Swahili case (7)-(9), none of the predicative possessive constructions of Chuvash involve a genuine verb. In (12), we have an existential predication, whereas example (13) is a zero-copula construction. Almost identical patterns are reported for the following sister-languages of Chuvash:

- Azerbaidjani (Landmann 2013, pp.7, 11)
- Kazakh (Landmann 2012, pp. 9, 14)
- Tatar (Landmann 2014b, pp. 13, 18)

In the cases of Bashkir (Ersen-Rasch 2009, pp. 34–35) and Turkish (Ersen-Rasch 2012, p. 60), the topic of lexB$_{poss}$-verb is not raised at all. However, based on the description of other predicative possessive categories in these languages, one might assume that the situation resembles that of Chuvash. Hungarian (Tompa 1972, pp. 104–106, 178, 192, 215) similarly lacks a lexB$_{poss}$-verb. Its B$_{poss}$-constructions generally have the form of S$_{poss}$-constructions with the possessor in the dative. The evidence for Komi (Beznosikova et al. 2003) and Mordvin (Ščankiva et al. 2011) is comparable to the Hungarian case.

For many languages of Eurasia and the (Trans-)Caucasus we also must argue ex negativo in the sense that the grammarians describe predicative possession in some detail without mentioning whether a lexB$_{poss}$-verb exists.

- Lezgian (Haspelmath 1993, pp. 311–326),
- Kalmyk (Benzing 1985, p. 56),
- Mari (Alhoniemi 1993, p. 50),

- Udmurt (Winkler 2011, p. 48–49).

On this basis, it is tempting to generalize other Daghestanian and Uralic languages of the European east. However, the linguistic sources consulted for these languages provide too little information on the systems of predicative possession to justify any claims of ours in this domain. Nevertheless, what can be said is that in all these languages lexH_{poss}-verbs do not exist. S_{poss}-constructions, on the other hand, are well established. Independent of the co-presence of a H_{poss}-verb, lexB_{poss}-verbs may thus be absent.

3.2 Indo-European Languages without Lexical BELONG-Verbs

lexB_{poss}-verbs are absent not only from the systems of predicative possession in many non-Indo-European languages of Europe. Several members of the Indo-European language family also either lack lexB_{poss}-verbs or it is at least doubtful that a verb of this kind exists. The situation is relatively uncontroversial in the cases of the Celtic language, Irish, and the Iranian language Kurdish (Kurmanci).

Irish has a lexical verb neither for H_{poss} nor for B_{poss}. The constructions, which are employed for predicative-possessive functions, involve PPs to identify the relation and mark the possessor. In (14), an existential construction with a spatial PP is used, which characterizes Irish H_{poss} to be construed on the Location Schema (Heine 1997, p. 51). In contrast, B_{poss} is expressed by a construction, which involves the copula and a PP headed by the preposition *le* 'with'. In this case, it is legitimate to speak of a realization of the Companion Schema, which is considered frequently to provide the basis of B_{poss}-constructions (Heine 1997, p. 57).

(14)   Irish: H_{poss} (Ó Siadhail 1985, p. 33)
       Tá   teach   ag   an   mbean.
       EXI   house   on   DEF   woman
       'The woman has a house.'

(15)   Irish: $B_{poss} = S_{poss}$ (Ó Siadhail 1985, p. 106)
       Is   le   Cáit   an   teach   seo.
       COP   with   Kate   DEF   house   DEM.PROX
       'This house belongs to Kate.'

The situation is almost identical in the two closest relatives of Irish, viz. Scots-Gaelic (Byrne 2004, pp. 71, 77) and Manx (Kewley Draskau 2008, pp. 46, 62, 181–182). As for Breton, the structural facts are very similar because there

is no LEXB$_{poss}$-verb. Its task is fulfilled by a S$_{poss}$-construction (Favereau 1997, pp. 218–219).

Kurdish poses more problems than Irish because, in contrast to the latter, the evidence of the absence of a LEXB$_{poss}$-verb is only indirect. According to Khan/Lescot (1986, pp. 177–179), there are two major types of H$_{poss}$-constructions both of which are basically existential predications. Apart from these instances of H$_{poss}$, the authors mention in passing that there is also a S$_{poss}$-construction, provided the possessee is interpreted as definite. The existence of a proper LEXB$_{poss}$-verb is not stated explicitly.

(16) Kurdish: H$_{poss}$ I (Khan/Lescot 1986, p. 177)
Hesp-ê   min   hebû.
Horse-IZAFE   1SG   EXI.3SG.PRET
'I had a horse.'

(17) Kurdish: H$_{poss}$ II (Khan/Lescot 1986, p. 178)
Min   hesp-ek   heye.
1SG   horse-INDEF   EXI.3SG.PRS
'I have a horse.'

(18) Kurdish: S$_{poss}$ (Khan/Lescot 1986, p. 178)
Hesp   yê   min   e.
horse   PRO   1SG   COP.3SG.PRS
'The horse is mine.'

The situation in Kurdish resembles that of Irish insofar as there is no evidence of either LEXH$_{poss}$-verbs or LEXB$_{poss}$-verbs. The constructions used in both languages for the expression of H$_{poss}$ involve existentials whereas that used for B$_{poss}$ contain the copula. It is very likely that the Kurdish scenario largely also applies in the case of the Iranian language Zaza, which is spoken in the vicinity of Kurdish populations in Anatolia. However, the reference grammars of Zaza that we had access to (Paul 1998; Selcan 1998) skip the issue of predicative possessive constructions completely. For the very same reason, we have also not been able to determine the structure of predicative possession in Ossetian (Hettich 2010) yet, although the little information that can be gathered from the grammatical sketch yields a picture, which greatly resembles that of Kurdish.

Independent of their genetic affiliation, the languages, which lack evidence of LEXB$_{poss}$-verbs, are located on the western and eastern outskirts of Europe. In the west, the absence of this kind of verb is sporadic. In contrast to these isolated

cases, languages without LEXB$_{poss}$-verbs seem to dominate in the European East. Thus, an areally meaningful pattern emerges if we study the distribution of the constructions of B$_{poss}$ in Europe.

4 European B$_{poss}$-Constructions as of Today: LEXB$_{poss}$-Verbs

In this section, we survey the synchronic distribution of the different cases of LEXB$_{poss}$-verbs in the languages of Europe. 47 of the 68 languages attest verbs of this kind—a figure which equals slightly more than 69 % of the entire sample.

Before we set out to discuss the data, a word of caution is called for. The English verb *belong* is polysemous in the sense that B$_{poss}$ is only one of its several readings. Even the prepositional verb *to belong to* allows for interpretations, which are not strictly possessive. According to Webster's (1994, p. 137), *belong to* has the following meanings:

- 'to be the property of ' (= B$_{poss}$-verb),
- 'to be a part or adjunct of ',
- 'to be a quality, function, or concern of '.

If a grammar or dictionary of a given language is written in English, it cannot be ruled out that references made therein to English *belong* connect the verb of the object language to one of the non-possessive meanings of the English verb.[2] Uncertainties of this kind can only be solved based on a sufficiently large corpus and/or a dedicated questionnaire.

Since the English verb is the namesake of the category under scrutiny here, we start with the discussion of B$_{poss}$ in English and its diffusion across neighboring languages (= Section 4.1). In Sections 4.2.-4.8., we trace the extension of the different isoglosses in Europe.

4.1 The English Case

The etymology of English *belong* is not yet entirely settled as the specialists disagree on several aspects of the verb's diachronic origin. In Harper's etymological online dictionary, for instance, we find the following entry for *belong*:

---

2   This is the case, e.g., with Scots-Gaelic *buin* 'belong to, pertain to, be related to' which often is confused with *bean* 'touch'. The very detailed illustration of the use of these verbs in Scots-Gaelic is clearly indicative of the absence of genuinely possessive meanings (Mark 2004, pp. 68, 102).

mid-14c., 'to go along with, properly relate to,' from *be-* intensive prefix, + *longen* 'to go,' from Old English *langian* 'pertain to, to go along with,' which is of uncertain origin but perhaps related to the root of *long* (adj.). Senses of 'be the property of' and 'be a member of' first recorded late 14c. Cognate with Middle Dutch *belanghen*, Dutch *belangen*, German *belangen*. Replaced earlier Old English *gelang*, with completive prefix *ge-*. (Harper: *Etymological online dictionary*, Web)

Klein (1966, p. 162) hypothesizes that the original meaning was 'to be alongside of'. Hoad (2003) assumes a connection to Old English *ġelang* 'at hand, be dependent on', whereas Collins's online dictionary (Web) suggests the development from Old English *langian* 'to belong' (!) via the intensive formation *be-* + *longen* 'to be suitable' to Middle English *bilangen*. Wood (1912) provides a long list of meanings for Old English *langian*, namely 'grow long; long for (= reach out for); summon; belong, pertinere'. These proposals are suggestive of a certain degree of insecurity as to the exact succession of meaning changes and related processes. A crucial question is, whether the B_{poss}-reading is a relatively late acquisition of Middle English or whether it dates back to the Old English period, provided the prefix-less verb already covered the same meaning as suggested by Collins. Since it is common practice in possession research to assume the original semantics of a given construction prominently featuring its main predicator to be decisive for the identification of the conceptual schemas that underlie the synchronic facts, providing an answer to our question is of some importance. Meanings like 'to go', 'to reach out', 'to summon' fall under the rubric of Heine's (1997, pp. 47–50) Action Schema, whereas 'to be alongside of' and 'to be suitable' represent static concepts of completely different kinds with the former relating to the Location Schema (Heine 1997, pp. 50–53). Some of the meanings invoke transitivity, whereas others clearly belong in the sphere of intransitivity such as 'to grow long'. We refrain from opting for any of the possibilities for lack of familiarity with the early stages of the diachrony of English.

In contrast to the misty prehistory of the English B_{poss}-verb, it can be shown that, once established in the possessive function, *belong* had some success in language contact. Nance (1978, pp. 16, 265) registers *longya* as an English loan-verb in his Cornish dictionary. The function of *longya* is that of a B_{poss}-verb in what is termed Late Cornish. Its use, however, is not recommended by the compiler of the dictionary who gives preference to copula-based constructions of the S_{poss}-type, etc. In Brown's (2001, pp. 200–204) grammar of Modern (= revitalized) Cornish, the chapter on possession mentions neither the English loan-verb nor B_{poss} at all.

In the next section, we put the focus on Romance B_{poss}-constructions not the least because they can be shown to influence those of differently affiliated neighboring languages.

4.2 The Romance Phylum and Its Outreach

The common Latin ancestry of the Romance languages also comes to the fore in the domain of B_{poss}. All members of this phylum, of which we are aware, employ LEXB_{poss}-verbs which are etymologically related to Latin *pertinere* 'to stretch/reach as far as, to be related to, to pertain to, to be suitable for', which, at least in the classical period, was devoid of properly possessive functions. The range of meanings of the Latin verb overlaps with that of the Old English verbs mentioned in the preceding section.

The modern forms of the B_{poss}-verb are as follows:

- Catalan *pertànyer à*
- French *appartenir à*
- Italian *appartenere a*
- Portuguese *pertencer a*
- Romanian *aparține*
- Spanish *pertenecer a*

Since all those Romance languages considered share the same B_{poss}-verb, it is likely that the possessive functions of *pertinere* (probably in the shape of *adpertinere*) date back to the Late or Vulgar Latin period. In the Romanian case, the B_{poss}-verb is assumed to be a direct loan from French, i.e., it entered the language relatively late in the 18th-19th century (Coteanu et al. 1975, p. 44). The French B_{poss}-verb is attested as early as the 11th century (Robert 1979, p. 84). In all of the above Romance languages, the B_{poss}-verb also has non-possessive meanings such as that of membership in a group, etc.

Furthermore, the six Romance languages are similar insofar as they all make use of a LEXH_{poss}-verb and give evidence of S_{poss}, too. For reasons of space, we illustrate this common system by way of discussing the examples (19)-(21) from French. The format of the S_{poss}-construction in (20) is unique to French within the Romance phylum—the other Romance languages reflect different patterns.

(19)   French: H_{poss}
       J'   ai   une   voiture.
       1SG   have.1SG   INDEF   car
       'I have a car.'

(20) French: S$_{poss}$
      La   voiture   est   à   moi.
      DEF   car   be.3SG   at   1SG.DAT
      'The car is mine/the car belongs to me.'

(21) French: B$_{poss}$
      La   voiture   m'   appartient.
      DEF   car   1SG.DAT   belong.3SG
      'The car is mine/the car belongs to me.'

At least two languages outside the Romance phylum display LEXB$_{poss}$-verbs, which are related directly or indirectly to the Romance equivalents. The Celtic language Welsh makes use of *perthyn(u)* 'to belong to, to be related, to pertain to, to form part of', whose first attestations were recorded as early as the 14th century (Thomas/Bevan/Donovan 2007, pp. 2781–2782). Whether it was integrated into Welsh via Middle English *perteine* (< Middle French *partenir*) or directly from Latin *pertinere*, is a question we cannot answer satisfactorily in this paper. Welsh is thus the sole Celtic language to boast a LEXB$_{poss}$-verb.

The case of the Afro-Asiatic language Maltese is as intriguing as the Welsh case. Maltese has borrowed the LEXB$_{poss}$-verb *appartiena* 'to belong, to be a member of' directly from Italian—a fact which is not surprising because about half of the Maltese lexicon has Romance origin. Originally, Maltese had no LEXH$_{poss}$-verb. The preposition *għand* 'at someone's place' has been grammaticalized for predicative possessive purposes so that it is nowadays classified as a (somewhat special) verb (termed pseudo-verb by Peterson 2009, pp. 199–200). There is also the expected S$_{poss}$-construction, cf. (22)-(24)[3].

(22) Maltese: H$_{poss}$ (Aquilina 1991, p. 967)
      Għand-i   ktieb   li   int   m'   għand-ek-x   bħal-u.
      have-1SG book   REL   2SG   NEG   have-2SG-NEG   like-3SG.M
      'I have got a book the like of which you have not got.'

---

3   Original English translations.

(23) Maltese: S$_{poss}$ (Aquilina 1991, p. 1395)
      Dal-ktieb   la   hu   tagħ-ha   u   lanqas   tagħ-hom.
      DEM:M:PROX:DEF-book   neither   3SG.M   of-3SG.F   and   nor   of-3PL
      'This book is neither hers nor theirs.'

(24) Maltese: B$_{poss}$ I (Aquilina 1987, p. 36)
      Dan   ir-raba'   j-appartieni   lill-familja   Cassia.
      DEM:M:PROX   DEF-field.COLL   3SG.M.IMPERF-belong   to:DEF-family   Cassia
      'This estate belongs to the Cassia family.'

This instance of a borrowed LEXB$_{poss}$-verb is distinct from the parallel Welsh case since there is also a partly synonymous Semitic LEXB$_{poss}$-verb, which competes with the Italian loan-verb., i.e., *għajjat* 'to shout' which can be employed in predicative possessive function as shown in (25).

(25) Maltese: B$_{poss}$ II (Aquilina 1991, p. 944)[4]
      ir-raba'   j-għajjat   lil-u
      DEF-field.COLL   3SG.M.IMPERF-shout   OBJ-3SG.M
      'The fields belong to him.'

In the absence of diachronic evidence of the age of both constructions in Maltese, we cannot determine whether *għajjat* chronologically precedes *appartiena* as a LEXB$_{poss}$-verb. If future research reveals that this hypothesis can be substantiated empirically, then we have a scenario in which the borrowed verb does not fill a gap in the replica language's system. Independent of the exact historical order of events, the facts discussed in this section speak in favor of an isogloss that embraces the entire Romance phylum, Welsh, and Maltese. Given the polysemy of the Latin etymon, it is again difficult to pin down the conceptual basis of the B$_{poss}$-constructions under review. A possible solution is to single out the first meaning of *pertinere* which is 'to stretch/reach as far as' and link it to the Old English meaning 'to reach out'. Under the condition that this very tentative suggestion holds true, we postulate an isogloss of ATTAINING, which presupposes Heine's Action Schema. Section 4.3 demonstrates that the isogloss reaches further into the Balkans.

---

4   Original English translation.

4.3 In the Balkans

Albanian and Greek are internal isolates of the Indo-European language family. They both are equipped with lexical verbs for H_{poss} and B_{poss}. The Greek LEXB$_{poss}$-verb *anḗkō* 'to belong' can be traced back to its ancestor *anḗkō* in Ancient Greek. In antiquity, this verb still had a much wider range of meanings, among which that of 'to reach as far (up) as' is prominently featured (Gemoll 1954, p. 69). Thus, the Greek LEXB$_{poss}$-verb ties in nicely with the above isogloss of ATTAINING.

Similarly, there is also a semantic bridge to link Albanian to the same isogloss. In (26)-(28), we give examples of typical instances of Albanian H_{poss}, S_{poss}, and B_{poss}.

(26)　Albanian: H_{poss} (Simoni 1978, p. 184)
　　　Kam　　　shumë　　punë.
　　　have:1SG　　much　　work
　　　'I have a lot of work.'

(27)　Albanian: S_{poss} (Buchholz/Fiedler 1987, p. 220)
　　　Ky　　　　　librër　është　　i　　　Agimit.
　　　DEM.M.PROX　book　be.3SG　GEN　Agim:GEN
　　　'This book is Agim's.'

(28)　Albanian: B_{poss} I (Simoni 1978, p. 158)
　　　libri　　　　më　　　përket　　　mua.
　　　book:DEF.M　1SG.DAT　belong:3SG　1SG.DAT
　　　'The book belongs to me.'

The LEXB$_{poss}$-verb *përkas* is not only equivalent to English *to belong to* but also has the (primary) meaning of 'to touch'. This action verb meaning connects Albanian to the isogloss as outlined in the previous section because TOUCHING and ATTAINING can be considered conceptual next-door neighbors of each other.

In Albanian, there is a second candidate for the status of a LEXB$_{poss}$-verb. The verb *takon* 'to meet, encounter, hit, befit, belong' can have possessive functions as shown in (29).

(29)　Albanian: B_{poss} II (Buchholz/Fiedler 1987, p. 479)
　　　Arti　　　　i　　takon　　　popullit.
　　　art:DEF.M　DAT　belong:3SG　people:DEF.DAT
　　　'Art belongs to the people.'

In addition to *takon* there is the polysemous verb *bie* 'to fall'. This verb is only marginally related to the domain of BELONGING because it can only be used in certain contexts to express BEFITTING. What makes this verb interesting for our study however is the fact that in the languages in the Albanian geographical neighborhood, *fall*-verbs have been functionalized as LEXB$_{poss}$-verbs.

In the South Slavic languages, the LEXB$_{poss}$-verbs consist of the basis *pad-* 'to fall' plus a prefix *pri-*. Bosnian, Croatian, Serbian, and Slovenian yield the same word-form, namely *pripadati*. Discounting the occasional detail, the Croatian examples in (30)-(32)[5] are representative of this group of four.

(30)   Croatian: H$_{poss}$ (Alexander 2006, p. 96)
Imaš   li   pri   sebi   novca?
have:2SG   Q   near   REFL:LOC   money:GEN
'Do you have any money on you?'

(31)   Croatian: S$_{poss}$ (Alexander 2006, p. 13)
Taj   pas   nije   moj.
DEM.M.DIST   dog   NEG:be.3SG   mine
'That dog isn't mine.'

(32)   Croatian: B$_{poss}$ (Alexander 2006, p. 98)
Ova   knjiga   pripada   učitelju.
DEM.F.PROX   book   belong:3SG   teacher:DAT
'This book belongs to the teacher.'

Macedonian has *pripad'a* with the same meaning. Bulgarian deviates from this pattern as it makes use of the Russism *prinadleža* (cf. below). However, it seems that *spada* had been in use as well as LEXB$_{poss}$-verb before being replaced by *prinadleža*. Beyond the Slavic sphere, we find similar cases in some of the Balkan varieties of Romani such as that of Prilep where the polysemous verb *perél* 'to fall' can also be employed as a LEXB$_{poss}$-verb (Boretzky/Igla 1994, p. 214).

What this section shows is that two isoglosses meet in the Balkans. As to the more southerly languages Albanian and Greek, their association with the ATTAINING-isogloss can hardly be denied. To the north of these languages, however, a new isogloss emerges, namely that of FALLING which comprises not only Slavic languages but also Indo-Aryan Romani and to some extent Albanian.

---

5   Original English translations.

4.4 Slavic—Part I: The Majority Option

In the majority of the Slavic languages spoken outside the Balkan region, the LEXB$_{poss}$-verbs are based on the positional verb meaning 'to lie (recline)' which combines with the spatial prefixes *na-* and *pri-*. In addition to the above-mentioned case of Bulgarian, we find the following examples:

- Russian: *pri-nad-ležat'*
- Belarusian: *na-ležac'*
- Ukrainian: *na-ležati*
- Kashubian: *nô-leżec ~ przë-nô-légac*
- Polish: *na-leżeć*
- Czech: *ná-ležet* (cf. below, Section 4.6)
- Slovak: *ná-ležať ~ pri-ná-ležať* (cf. below, Section 4.6)

Except for Russian, all these Slavic languages have full-blown LEXH$_{poss}$-verbs. The examples (33)-(35) reflect the Ukrainian situation.

(33) Ukrainian: H$_{poss}$

    a. Vona maje malen'kyj sad.
       she:NOM have:3SG.PRS small:ACC garden:ACC
       'She has a small garden.'

    b. V neji je malen'kyj sad.
       in she:LOC be:3SG.PRS small:ACC garden:NOM
       'She has a garden.'

(34)   Ukrainian: S$_{poss}$
    Cej sad jiji.
    this garden she:GEN
    'This garden is hers.'

(35)   Ukrainian: B$_{poss}$
    Cej sad naležyt' jij.
    this garden belong:3SG she:DAT
    'This garden belongs to her.'

The Slavic data of this section constitute a third isogloss. This isogloss can be labeled the LYING-isogloss. It is interesting to see that this isogloss cuts across the Slavic phylum to divide it into a southern and a northern branch. In the south, the FALLING-isogloss dominates whereas in the north the predominant isogloss

is that of LYING. Bulgarian (cf. Section 4.3) as well as Czech and Slovak attest to the coexistence of several options as discussed in Section 4.6.

For Russian, for instance, Tabačenko (2011) and Vinogradov (1994) assume that *prinadležat'* has been calqued on foreign patterns. The two authors disagree as to the languages, which have served as a model: Greek, Latin, and/or German. Nevertheless, our sources concur with each other insofar as both consider the late 18th century the period in which the possessive reading of the erstwhile positional verb emerged (most probably first in the administrative style). Non-possessive meanings of the Russian LEXB$_{poss}$-verb became obsolete in the beginning of the 19th century.

As will result from the next section, the FALLING-LYING divide is not the only factor responsible for the internal diversity of the Slavic phylum in the domain under review. Moreover, there is strong evidence of language contact with Germanic languages being the cause of further diversification. To prove this hypothesis, we must take a lengthy detour by way of looking at the Germanic data first.

4.5 Germanic

Excluding English, all members of the Germanic phylum boast LEXB$_{poss}$-verbs, which are based on verbs with the original meaning 'to hear'. In the following list of Germanic LEXB$_{poss}$-verbs we underline the shared cognate:

- Danish : *til<u>høre</u>*
- Dutch: *(be)<u>horen</u>*
- Faroese: *<u>hoyra</u> til*
- Frisian: *(be)<u>hearre</u>*
- German: *ge<u>hören</u>*
- Icelandic: *til<u>heyra</u>*
- Low German: *(to)<u>höören</u>*
- Luxembourgish: *ge<u>héieren</u>*
- Norwegian: *til<u>høre</u>*
- Swedish: *till<u>höra</u>*

This is evidence of a surprisingly homogeneous behavior of the members of this phylum, which largely can compete with that reflected by the Romance phylum. *Mutatis mutandis*, the German examples in (36)-(38) are representative of the languages enumerated above.

(36)   German: H_{poss}
        Ich   habe   ein   Buch.
        1SG   have:1SG   INDEF.ACC   book
        'I have a book.'

(37)   German: S_{poss}
        Das   Buch   ist   meins.
        DEF.N   book   be.3SG   POR.1SG:N
        'The book is mine/belongs to me.'

(38)   German: B_{poss}
        Das   Buch   gehört   mir.
        DEF.N   book   belong:3SG   1SG.DAT
        'The book belongs to me.'

According to the DWDS (Web), *gehören* originally was an intensive formation of the perception verb *hören* 'to hear' and acquired its possessive meanings only in the 14th century, whereas the original meaning was subsequently lost. Conceivably, the derivative of a verb of hearing with possessive function has diffused via (Middle) Low German into mainland Scandinavia and from there to the Faroe Islands and Iceland to yield a solid block of languages, which partake in the HEARING-isogloss. This isogloss, however, is not restricted to the Germanic phylum. The next section provides evidence of the spread of the Germanic pattern into the territory of Slavic, Baltic, and Uralic languages.

4.6   Slavic—Part II: Germanic Influence and Parallels in Baltic and Uralic Languages

As of now, we know of nine languages spoken along the eastern borderline of the territory occupied by Germanic languages in Europe, which employ $LEXB_{poss}$-verbs whose etymological connection to verbs of hearing is more or less evident. In Tab. 1, we present these languages from North to South with their verbs of hearing and the corresponding $LEXB_{poss}$-verb. Grey shading is used to highlight the Czech case because the verb of hearing and that of belonging are not direct cognates of each other. Nevertheless, they seem to correspond to the patterns exemplified by the other languages in Tab. 1.

In Saami, the two verbs are homophonous. In contrast to this case of identity, the segmental chains of the verb of hearing and the $LEXB_{poss}$-verb differ at least slightly in each of the other languages. These differences notwithstanding,

Tab. 1: Eastern Central European languages that partake in the HEARING-isogloss

language	hear/listen	belong
Saami	*gullet*	*gullet*
Finnish	*kuulla*	*kuulua*
Estonian	*kuulma*	*kuuluma*
Lithuanian	*klausyti*	*priklausyti*
Lower Sorbian	*słuchaś*	*słušaś*
Upper Sorbian	*słuchać*	*słušeć*
Czech	*poslouchat*	*příslušet*
Slovak	*poslúchať*	*prislúchať*
Burgenland Croatian	*slušati*	*slišiti* (also: *pripadati*)

the similarities are evident. In several of these languages all of which have been exposed to language contact with members of the Germanic phylum for an extended period of time. Since eastern relatives of the languages in Tab. 1 do not share this property, we obviously face a case of diffusion of a Germanic pattern into the territory of the neighboring languages of different genetic affiliation.

In the Finnish etymological dictionary, Hekkinen (2007, pp. 523–524) ponders the idea that the semantic association of hearing with belonging might be influenced by the relevant patterns in Swedish and German. The author also assumes that the possessive function of the erstwhile perception verb emerged in the 18th century. We assume that this explanation also holds for Estonian and Saami (in the latter case perhaps via the mediation of Finnish itself). In Finnish, Estonian, and Lithuanian, the LEXB$_{poss}$-verb is more complex than the corresponding verb of hearing since there is an additional syllable in the former. The higher complexity has a morphological explanation in the Lithuanian case because the LEXB$_{poss}$-verb is derived from the verb of hearing by way of prefixing *pri-* to *klausyti*. In the two Balto-Finnic languages there is an additional vowel *-u-* which possibly reflects an intensifying formation, i.e., a derivation on the basis of the verb of hearing. For the two Sorbian varieties however, it is impossible to pin down a derivational relation based on segmental complexity. For the time being and in analogy to the previous cases, we assume that the LEXB$_{poss}$-verbs go back to intensive verbs derived from the perception verb.

The Finnish examples in (39)-(41) confirm once more that a language without a LEXH$_{poss}$-verb may host a LEXB$_{poss}$-verb.

(39)   Finnish: H$_{poss}$ (Hirvensalo 1975, p. 508)
        Häne-llä   on   paljon   rahaa.
        3SG-ADESS   be.3SG   much   money:PTV
        'He has a lot of money.'

(40)   Finnish: S$_{poss}$ (Hirvensalo 1975, p. 438)
        Puutarha   on   hän-en.
        garden   be.3SG   3SG-GEN
        'The garden is his/hers.'

(41)   Finnish: B$_{poss}$ (Hirvensalo 1975, p. 438)
        Puutarha   kuuluuu   häne-lle.
        garden   belong:3SG   3SG-ALL
        'The garden belongs to him/her.'

Slovak is special because it is reported to have three different LEXB$_{poss}$-verbs, one of which is *náležať ~ prináležať* mentioned in Section 4.4. This verb connects Slovak to the majority of the Slavic languages north of the Balkans. This verb is said to be typical of literary Slovak. The same stylistic classification applies to the synonymous *prislúchať* in Tab. 1. *Patriť*, however, is more commonly used, which formerly meant 'to look', i.e., another verb of perception has been transformed into a LEXB$_{poss}$-verb. Czech provides an almost perfect parallel to the Slovak case. According to Agnes Kim (personal communication), Czech makes use of three LEXB$_{poss}$-verbs: *náležet*, *patřit* and *přislušet*.[6]

To summarize the above discussion, we state that the likeliest center of diffusion of the HEARING-isogloss is located in the Germanic phylum. The Germanic pattern has been calqued by languages of Eastern Central Europe, which belong to different language families and phyla. For some of the languages, the integration of the Germanic pattern into their system of predicative possession seems to be a relatively recent phenomenon, whereas their LEXB$_{poss}$-verbs display morphological properties, which must predate the possessive reading by centuries.

---

6   The Czech etymological dictionary (Rejzek 2015, p. 501) argues that the word *patřit* was already attested in the 14th century with largely unclear origins. One possible connection is with Proto-Indo-European **peh$_2$-tro* 'fodder'. The meaning change from 'to nourish' to 'to belong' presupposes intermediate meanings such as 'to look at something' or 'to be in one's range of vision'. As similar semantic change is assumed for *příslušet,* which is a derivation from *slušet* 'to fit'.

4.7 Potential Loners

Three languages need to be mentioned very briefly before we can present our conclusions, namely Latvian, Armenian, and Abkhaz. Lithuanian's sister language Latvian partakes neither in the HEARING-isogloss nor in the LYING-isogloss and thus behaves differently from its neighbors and relatives. Its LEXB$_{poss}$-verb *piederēt* is a prefixal derivation from *derēt* 'to befit, to be appropriate'.[7] Since Latvian at the same time lacks a LEXH$_{poss}$-verb, we have further evidence of the relative dissociation of the two predicative possessive categories because the constructions, which express H$_{poss}$ and B$_{poss}$, realize different conceptual schemas—in this case Heine's Goal Schema for the former and, in the case of B$_{poss,}$ something that has not yet been labeled (because being appropriate or befitting can hardly accommodate the Action Schema), cf. (42)-(43).

(42)   Latvian: H$_{poss}$ (Smiltniece 2015, p. 351)
        Meitenei   ir   grāmata
        girl:DAT   be.3   book
        'The girl has a book.'

(43)   Latvian: B$_{poss}$ (Smiltniece 2015, p. 351)
        Māja   pieder   brālim
        house   belong.3   brother:DAT
        'The house belongs to the brother.'

In contrast to Latvian, Armenian has lexical verbs for both H$_{poss}$ and B$_{poss}$. The LEXH$_{poss}$-verb is already attested on the earliest documented stages of the language. As to *patkanel* 'to belong', the documentation starts with non-possessive meanings such as 'be appropriate' in the 5th century, whereas its possessive functions come to the fore in the 10th century. Armenian thus sides with Latvian as representative of the BEFITTING-type, which does not yield a geographically connected isogloss on the map. The further etymology of the Armenian LEXB$_{poss}$-verb is somewhat dubious. What is clear is that it is a loan-verb borrowed from an Iranian donor-language (presumably Parthian). The original meaning can only be left to speculation. Andrea Scala (personal communication) suggests a connection to Old Indic *pati* 'owner'. Sentential examples of both Armenian verbs are given in (44)-(45)[8].

---

7   The editors suggest that the Latvian derivation might be a parallel to the secondary meanings of the LexB$_{poss}$-verbs in Czech, and especially, to *slušet* 'to fit' (cf. Footnote 8).
8   Original English translations.

(44)   Armenian: H$_{poss}$ (Dum-Tragut 2009, p. 65)
        Ērex-ek'   č'-un-eink'
        child-NOM.PL   NEG-have-PRET.1PL
        'We had no children.'

(45)   Armenian: B$_{poss}$ (Dum-Tragut 2009, p. 90)
        Alek'sandr-i-n   patkan-el   ē   bnakaran-i   ¾-ĕ
        Alexander-DAT-DEF belong-PTCP.PERF be.3SG apartment-DAT ¾-DEF
        'Three quarters of the apartment belonged to Alexander.'

Similarly, we are still struggling with the same etymological problem as Stassen (2009, p. 295), who claimed that the origin of the Abkhaz LEXH$_{poss}$-verb *-ma-* "is unknown". In our case, this problem also applies to the case of the LEXB$_{poss}$-verb *t^{o}'ə*, cf. (46)-(47)[9].

(46)   Abkhaz: H$_{poss}$ (Hewitt 1989, p. 82)
        a-pàra   Ø-sə̀-ma-cəpx̣aʒa
        DEF-money   3SG-1SG-have-every_time
        'Every time I had money…'

(47)   Abkhaz: B$_{poss}$ (Hewitt 1989, p. 64)
        a-pħ^{o}ə̀s   yə̀-l-t^{o}'ə-w
        DEF-woman   REL-3SG.F.DAT-belong-NFIN.STAT.PRS
        'the woman's [= that which belongs to the woman]'

According to Nana Machavariani (personal communication), the Abkhaz LEXB$_{poss}$-Verb is etymologically connected to a positional verb 'to sit'. In the absence of further information, we can only state that this Northwest Caucasian language sides with (genetically unrelated) Georgian (cf. (3)-(4) above) and Armenian (cf. (44)-(45) above) insofar as it employs lexical verbs for the two predicative possessive categories whereas other languages in the eastern outskirts of the continent lack evidence of LEXH$_{poss}$-verbs as well as LEXB$_{poss}$-verbs. The Georgian LEXB$_{poss}$-verb is etymologically associated with a verb which has the meaning of 'to cut off (a piece) for somebody' (Winfried Böder, personal communication). The Caucasus thus seem to be particularly diverse as to the conceptual basis of the expressions of B$_{poss}$ in this area. None of the three

---

9   Original English translations.

languages from this region are involved in one of the major isoglosses as identified elsewhere in Europe.

5 Conclusions

In this study, we have identified five major isoglosses, which serve as geo-linguistic subdivisions of the European linguistic map. First, the contrast between the presence and absence of LEXB$_{poss}$-verbs yields a relatively clear areal linguistic picture. Except for the Caucasian region, on the western and eastern fringes of the continent, LEXB$_{poss}$-verbs are absent, whereas verbs of this kind abound in the large area between the two margins. Furthermore, those languages which boast LEXB$_{poss}$-verbs, give evidence of different conceptual sources as identified etymologically. There are four isoglosses which can be identified on the basis of the original meaning of the LEXB$_{poss}$-verb (discounting idiosyncratic cases), namely

- the ATTAINING-isogloss which stretches from Britain southwards and extends into the Balkans uniting the entire Romance phylum with English, Welsh, Albanian, and Greek;
- the FALLING-isogloss, which is primarily used in the Balkans, involving the South Slavonic languages and local varieties of Romani as well as (possibly) also Albanian with its secondary options;
- the LYING-isogloss in which most of the Slavic languages north of the Balkans partake;
- the HEARING-isogloss hosts the Germanic languages (except English) and has made inroads into the territories of the immediate Uralic, Baltic, and Slavic neighbors east of the language boundary.

The latter isogloss is a piece of evidence of the linguistic interaction in Eastern Central Europe suggesting that language contact is largely responsible for some of the peculiarities of languages of this zone, which sets them apart from their next-of-kin to the east and south. Only a small number of languages with LEXB$_{poss}$-verbs (Abkhaz, Armenian, Georgian, and Latvian) cannot be assigned to any of the above isoglosses.

For all four of the isoglosses, and some of the isolated cases, the majority of the attested LEXB$_{poss}$-verb patterns reflects a binary internal structure [PREFIX-V$_{non-possessive}$]$_{belong}$ such as Albanian *për*$_{prefix}$-*kas*, English *be*$_{prefix}$-*long*, Swedish *til*$_{prefix}$-*höra*, Latvian *pie*$_{prefix}$-*derēt*, Croatian *pri*$_{prefix}$-*padati*, etc. This means that the LEXB$_{poss}$-verbs are mostly derived from verbs with non-possessive meanings, i.e., the B$_{poss}$-function is clearly secondary. Note that formally derived LEXH$_{poss}$-verbs are not attested in our sample. Within the HEARING-isogloss, Lithuanian is the

only non-Germanic language to attest to the above binary structure. The $\text{lexB}_{\text{poss}}$-verbs of the other non-Germanic languages of this isogloss are different—with stem extensions and internal modification, so that a relatively remote genesis of these $\text{lexB}_{\text{poss}}$-verbs is possible.

Our data additionally show that $\text{lexB}_{\text{poss}}$-verbs are compatible with both the presence and absence of $\text{lexH}_{\text{poss}}$-verbs in a given language. There are 40 languages, which boast a $\text{lexB}_{\text{poss}}$-verb alongside a $\text{lexH}_{\text{poss}}$-verb. Thus, 85 % of all those languages which are equipped with a $\text{lexB}_{\text{poss}}$-verb, also have a $\text{lexH}_{\text{poss}}$-verb. In seven languages, the $\text{lexB}_{\text{poss}}$-verb lacks a corresponding $\text{lexH}_{\text{poss}}$-verb. In contrast, $\text{lexH}_{\text{poss}}$-verbs are present in a given language preferably, if the language also has a $\text{lexB}_{\text{poss}}$-verb. Basque is the only example of a violation of this preference. It is striking that, with 47 cases, $\text{lexB}_{\text{poss}}$-verbs are more widely attested in Europe than $\text{lexH}_{\text{poss}}$-verbs, which for only 41 cases. 20 languages without $\text{lexB}_{\text{poss}}$-verbs also lack a $\text{lexH}_{\text{poss}}$-verb.

For research into linguistic possession in general, this means that the two predicative possessive notions constitute a relatively tight-knit paradigm. H_{poss} and B_{poss} seem to be relatively strongly connected to each other although we have not come across a single instance of a $\text{lexB}_{\text{poss}}$-verb being derived from the corresponding $\text{lexH}_{\text{poss}}$-verb or vice versa. If it is true that B_{poss} is relatively closely associated with ownership, whereas H_{poss} has a much wider range of conceptual associations, then it makes sense to compare B_{poss} to a third category of predicative possession as well, viz equivalents of the English verb *to own* in other languages or O_{poss}-constructions. Only if all three of these notions are investigated as a network of categories, will we be able to fully understand how the grammar of predicative possession is organized.

Before we can achieve this, we need to refine our methodology and enlarge our database. The limitations of this study mainly result from our decision to take the information given in the descriptive linguistic sources at face value to the exclusion of all other kinds of sources. In the future, we will have to rely on corpus data to a larger degree and include questionnaire-based findings as well. To complement the qualitative side of our project it is also necessary to look at the data quantitatively. We already know (Stolz/Levkovych accepted) that having a $\text{lexB}_{\text{poss}}$-verb does not mean the same for any two languages since the range of contexts in which these verbs are employed may vary considerably from one language to the other.

Acknowledgments

Abbreviations

1/2/3 = 1st/2nd/3rd person, ABS = absolutive, ACC = accusative, ADESS = adessive, ALL = allative, B_{poss} = *belong*-possession, CHV = character vowel, CL = class, COLL = collective, CONN = connector, COP = copula, DAT = dative, DEF = definite, DEM = demonstrative, DIST = distal, ERG = ergative, EXI = existential, F = feminine, GEN = genitive, H_{poss} = *have*-possession, IMPERF = imperfective, INDEF = indefinite, IZAFE = izafe-marker, LEX = lexical, LOC = locative, M = masculine, NEG = negation, NFIN = nonfinite, NOM = nominative, N = neuter, OBJ = object, O_{poss} = *own*-possession, PERF = perfect, PL = plural, POR = possessor, POSS = possessive, PRET = preterit, PRO = pronoun, PRS = present tense, PROX = proximal, PTCP = participle, PTV = partitive, Q = question marker, REFL = reflexive, REL = relative, S_{poss} = substantival possession, SG = singular, STAT = stative, SBJ = subject

Appendix I: Map of B$_{poss}$-Isoglosses in Europe

[Languages with LExB$_{poss}$-verbs, which cannot be associated with one of the isoglosses, are marked in gray shading. On the map, the individual languages are represented by codes. The codes are disclosed in Appendix II]
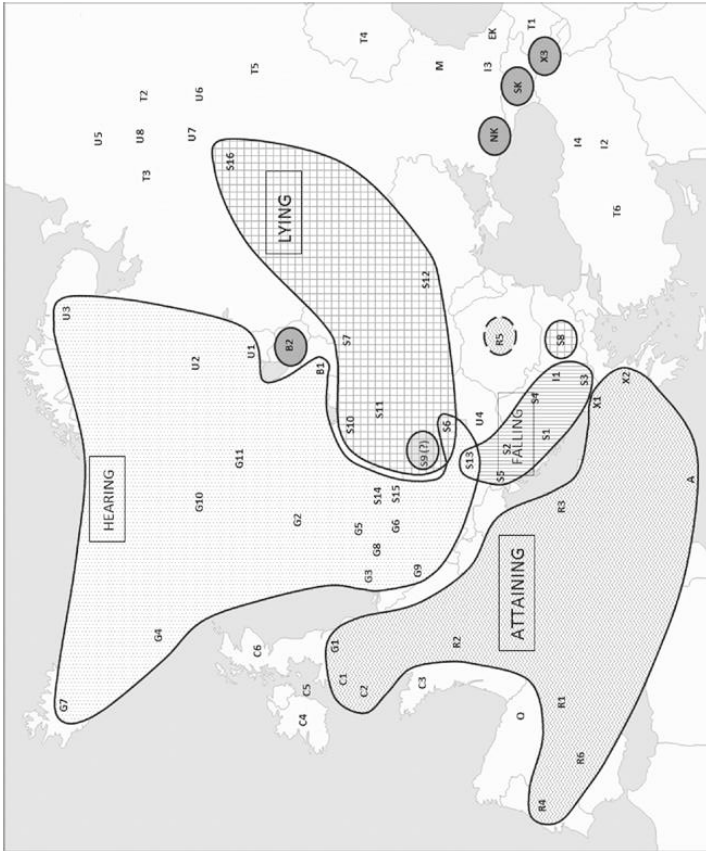
Appendix II: European Languages Represented on the Map

[The languages are ordered top-down as follows: The upper part of the table is dedicated to those languages for which the use of lexical verbs for the expression of both H_{poss} and B_{poss} is attested. Languages, which belong to the same isogloss and have identical genetic affiliation, are ordered alphabetically. The bottom part of the table is reserved for languages without a lexical verb in predicative possessive function. Unclear cases are identified by a question mark.]

language	LEXH$_{poss}$	LEXB$_{poss}$	isogloss	affiliation	code
English	yes	yes	ATTAINING	Germanic	G1
Catalan	yes	yes	ATTAINING	Romance	R1
French	yes	yes	ATTAINING	Romance	R2
Italian	yes	yes	ATTAINING	Romance	R3
Portuguese	yes	yes	ATTAINING	Romance	R4
Romanian	yes	yes	ATTAINING	Romance	R5
Spanish	yes	yes	ATTAINING	Romance	R6
Albanian	yes	yes	ATTAINING	Albanian	X1
Greek	yes	yes	ATTAINING	Greek	X2
Romani	yes (Prilep)	yes	FALLING	Indo-Aryan	I1
Bosnian	yes	yes	FALLING	Slavic	S1
Croatian	yes	yes	FALLING	Slavic	S2
Macedonian	yes	yes	FALLING	Slavic	S3
Serbian	yes	yes	FALLING	Slavic	S4
Slovenian	yes	yes	FALLING	Slavic	S5
Slovak	yes	yes	LYING/ HEARING/ (LOOKING)	Slavic	S6
Belarusian	yes	yes	LYING	Slavic	S7
Bulgarian	yes	yes	LYING	Slavic	S8
Czech	yes	yes	LYING / LOOKING / HEARING (?) (BEFITTING)	Slavic	S9
Kashubian	yes	yes	LYING	Slavic	S10
Polish	yes	yes	LYING	Slavic	S11
Ukrainian	yes	yes	LYING	Slavic	S12
Danish	yes	yes	HEARING	Germanic	G2
Dutch	yes	yes	HEARING	Germanic	G3

language	LEXH$_{poss}$	LEXB$_{poss}$	isogloss	affiliation	code
Faroese	yes	yes	HEARING	Germanic	G4
Frisian	yes	yes	HEARING	Germanic	G5
German	yes	yes	HEARING	Germanic	G6
Icelandic	yes	yes	HEARING	Germanic	G7
Low German	yes	yes	HEARING	Germanic	G8
Luxembourgish	yes	yes	HEARING	Germanic	G9
Norwegian	yes	yes	HEARING	Germanic	G10
Swedish	yes	yes	HEARING	Germanic	G11
Burgenland Croatian	yes	yes	FALLING/ HEARING	Slavic	S13
Lower Sorbian	yes	yes	HEARING	Slavic	S14
Upper Sorbian	yes	yes	HEARING	Slavic	S15
Lithuanian	yes	yes	HEARING	Baltic	B1
Armenian	yes	yes	BEFITTING	Armenian	X3
Abkhaz	yes	yes	(SITTING)	Northwest Caucasian	NK
Georgian	yes	yes	(CUTTING OFF)	Kartvelian	SK
Latvian	no	yes	BEFITTING	Baltic	B2
Russian	no	yes	LYING	Slavic	S16
Estonian	no	yes	HEARING	Uralic	U1
Finnish	no	yes	HEARING	Uralic	U2
Saami	no	yes	HEARING	Uralic	U3
Maltese	(no)	yes	ATTAINING	Afro-Asiatic	A
Welsh	no	yes	ATTAINING	Celtic	C1
Cornish	no	(yes)	ATTAINING	Celtic	C2
Basque	yes	no		isolate	Q
Azerbaidjanian	no	no		Turkic	T1
Bashkir	no	no		Turkic	T2
Chuvash	no	no		Turkic	T3
Kazakh	no	no		Turkic	T4
Tatar	no	no		Turkic	T5
Turkish	no	no		Turkic	T6
Hungarian	no	no		Uralic	U4
Komi	no	no		Uralic	U5
Mordvin	no	no		Uralic	U6
Breton	no	no		Celtic	C3

language	LEXH$_{poss}$	LEXB$_{poss}$	isogloss	affiliation	code
Irish	no	no		Celtic	C4
Manx	no	no		Celtic	C5
Scots-Gaelic	no	no		Celtic	C6
Kurdish	no	no		Iranian	I2
Ossetian	no	?		Iranian	I3
Zaza	no	?		Iranian	I4
Mari	no	?		Uralic	U7
Udmurt	no	?		Uralic	U8
Kalmyk	no	?		Mongolian	M
Lezgian	no	?		Daghestanian	EK

References

Aikhenvald, Alexandra Y.: "Possession and ownership. A cross-linguistic perspective". In: Aikhenvald, Alexandra Y./Dixon, R.M.W. (eds.): *Possession and Ownership. A Cross-Linguistic Typology* (= Explorations in Linguistic Typology). Oxford: Oxford University Press 2013, pp. 1–64.

Alexander, Ronelle: *Bosnian, Croatian, Serbian. A Grammar with Sociolinguistic Commentary*. Madison: The University of Wisconsin Press 2006.

Alhoniemi, Alho: *Grammatik des Tscheremissischen (Mari)*. Hamburg: Buske 1993.

Aquilina, Joseph: *Maltese-English Dictionary. Volume Two: M-Z and Addenda*. Malta: Midsea 1991.

Benzing, Johannes: *Kalmückische Grammatik zum Nachschlagen* (= Turkologica 1). Wiesbaden: Harrassowitz 1985.

Beznosikova, L. M. et al.: *Russko-komi slovar'*. Syktyvkar: Komi knižnoe izdatel'stvo 2003.

Boretzky, Norbert/Igla, Birgit: *Wörterbuch Romani – Deutsch – Englisch. Mit einer Grammatik der Dialektvarianten*. Wiesbaden: Harrassowitz 1994.

Bril, Isabelle: "Ownership, part-whole, and other possessive-associative relations in Nêlêmwa (New Caledonia)". In: Aikhenvald, Alexandra Y./Dixon, R.M.W. (eds.): *Possession and Ownership. A Cross-Linguistic Typology* (= Explorations in Linguistic Typology). Oxford: Oxford University Press 2013, pp. 65–89.

Brown, Wella: *A Grammar of Modern Cornish*. Callington: The Cornish Language Board 2001.

Buchholz, Oda/Fiedler, Wilfried: *Albanische Grammatik*. Leipzig: Enzyklopädie 1987.

Byrne, Michel: *Gràmar na Gàidhlig*. Cearsiadar: Stòrlann-Acair 2004.

Clark, Eve: "Locationals: existential, locative and possessive constructions". In: Greenberg, Joseph H. (ed.): *Universals of Human Language. Volume 4: Syntax*. Stanford: Stanford University Press 1978, pp. 85–126.

Collins: *Free Online Dictionary*, retrieved 27. 12. 2017, from https://www.collinsdictionary.com/dictionary/english/belong.

Coteanu, Ion et al.: *Dicționarul explicativ al limbii române*. București: Academia Republicii Socialiste România 1975.

Dixon, R.M.W.: *Basic Linguistic Theory. Volume 3: Further Grammatical Topics*. Oxford: Oxford University Press 2012.

Dum-Tragut, Jasmine: *Armenian* (= London Oriental and African Language Library 14). Amsterdam, Philadelphia: Benjamins 2009.

DWDS = Berlin-Brandenburgische Akademie der Wissenschaften: *DWDS = Digitales Wörterbuch der deutschen Sprache*, retrieved 27. 12. 2017, from https://www.dwds.de/.

Ersen-Rasch, Margarete I.: *Baschkirisch. Lehrbuch für Anfänger und Fortgeschrittene*. Wiesbaden: Harrassowitz 2009.

Ersen-Rasch, Margarete I.: *Türkische Grammatik*. Wiesbaden: Harrassowitz 2012.

Fähnrich, Heinz: *Kurze Grammatik der georgischen Sprache*. Leipzig: Enzyklopädie 1986.

Favereau, Francis: *Grammaire du breton contemporain*. Morlaix: Skol Vreizh 1997.

Gemoll, Wilhelm: *Griechisch-deutsches Schul- und Handwörterbuch*. München, Wien: Freytag & Hölder – Pichler – Tempsky 1954.

Harper, Douglas: *Online Etymology Dictionary*, retrieved 27. 12. 2017, from https://www.etymonline.com/.

Haspelmath, Martin: *A Grammar of Lezgian* (= Mouton Grammar Library 9). Berlin, New York: Mouton de Guyter 1993.

Haspelmath, Martin: "Explaining alienability contrasts in adpossessive constructions: predictability vs iconicity". *Zeitschrift für Sprachwissenschaft* 36 (2), 2017, pp. 193–232.

Heine, Bernd: *Possession. Cognitive sources, forces, and grammaticalization* (= Cambridge Studies in Linguistics 83). Cambridge: Cambridge University Press 1997.

Hekkinen, Kaisa: *Nykysuomen etymologinen sanakirja*. Juva: Bookwell 2007.

Hettich, Bela: *Ossetian* (= Languages of the World 475). München, Newcastle: LINCOM Europe 2010.

Hewitt, B. George: "Abkhaz". In: Hewitt, B. George (ed.): *The Indigenous Languages of the Caucasus. Volume 2: The North West Caucasian Languages*. Delmar, NY: Caravan 1989, pp. 38–88.

Hirvensalo, Lauri: *Saksalais-suomalaisen suursanakirja*. Porvoo, Helsinki: Söderström 1975.

Hoad, T. F.: *The Concise Oxford Dictionary of English Etymology*. Oxford: Oxford University Press 2003.

Hualde, José Ignacio/Ortiz de Urbina, Jon: *A Grammar of Basque* (= Mouton Grammar Library 26). Berlin, New York: Mouton de Gruyter 2003.

Kewley Draskau, Jennifer: *Practical Manx*. Liverpool: Liverpool University Press 2008.

Khan, Emir Djeladet Bedir/Lescot, Roger: *Kurdische Grammatik*. Bonn: Verlag für Kultur und Wissenschaft 1986.

Khizanishvili, Tamar: "Possession im Georgischen". In: Stolz, Thomas/Stroh, Cornelia (eds.): *Possession, Quantitative Typologie und Semiotik. Georgisch, Irisch, Türkisch* (= Diversitas Linguarum 11). Bochum: Brockmeyer 2006, pp. 1–76.

Klein, Ernst: *Comprehensive Etymological Dictionary of the English Language*. Amsterdam, etc.: Elsevier 1966.

König, Ekkehard/Haspelmath, Martin: "Der europäische Sprachbund". In: Reiter, Norbert (ed.): *Eurolinguistik. Ein Schritt in die Zukunft*. Wiesbaden: Harrassowitz 1999, pp. 111–128.

Kwon, Joung-Mi: *Possession im Swahili* (= Afrikanistische Monographien 7). Köln: Institut für Afrikanistik 1995.

Lafitte, Pierre : *Grammaire basque (navarro-labourdin littéraire)*. Donostia: Elkarlanean 1998.

Landmann, Angelika: *Kasachisch. Kurzgrammatik*. Wiesbaden: Harrassowitz 2012.

Landmann, Angelika: *Aserbaidschanisch. Kurzgrammatik*. Wiesbaden: Harrassowitz 2013.

Landmann, Angelika: *Tschuwaschisch. Kurzgrammatik*. Wiesbaden: Harrassowitz 2014a.

Landmann, Angelika: *Tatarisch. Kurzgrammatik*. Wiesbaden: Harrassowitz 2014b.

Levkovych, Nataliya/Mazzitelli, Lidia Federica/Stolz, Thomas: "Slavic vs. SAE. An areal-linguistic comparison on two different levels". In: Danylenko,

Andrij/Nomachi, Motoki (eds.): *Slavic in the language map of Europe*. Berlin, Boston: De Gruyter Mouton, to appear (accepted).

Mark, Colin: *The Gaelic-English Dictionary*. London, New York: Routledge 2004.

Mazzitelli, Lidia Federica: *The Expression of Predicative Possession. A Comparative Study of Belarusian and Lithuanian* (= Studia Typologica 15). Berlin, Boston: De Gruyter Mouton 2015.

Mazzitelli, Lidia Federica: "Predicative possession in the languages of the Circum-Baltic area". *Folia Linguistica* 51 (1), 2017, pp. 1–60.

Nance, R. Morton: *An English-Cornish and Cornish-English Dictionary*. Penzance: The Cornish Language Board 1978.

Ó Siadhail, Mícheál: *Lehrbuch der irischen Sprache*. Hamburg: Buske 1985.

Paul, Ludwig: *Zazaki. Grammatik und Versuch einer Dialektologie* (= Beiträge zur Iranistik 18). Wiesbaden: Reichert 1998.

Peterson, John: "Pseudo-verbs: an analysis of non-verbal (co-)predication in Maltese". In: Comrie, Bernard et al. (eds.): *Introducing Maltese Linguistics* (= Studies in Language Companion Series 113). Amsterdam, Philadelphia: Benjamins 2009, pp. 181–204.

Rejzek, Jiří : *Český etymilogický slovník*. Praha : Leda 2015.

Robert, Paul: *Le petit Robert. Dictionnaire alphabétique et analogique de la langue française*. Paris: Société du Nouveau Littré 1979.

Ščankina, V. I. et al.: *Russko-mokšansko-erzjanskij slovar'*. Saransk: Povolžskij centr kul'tur finno-ugorskich narodov 2011.

Seiler, Hansjakob: "Zum Problem der sprachlichen Possessivität". *Folia Linguistica* 6, 1973, pp. 231–250.

Seiler, Hansjakob: *Possession as an Operational Dimension of Language* (= Language Universals Series 2). Tübingen: Narr 1983.

Selcan, Zülfü: *Grammatik der Zaza-Sprache. Nord-Dialekt (Dersim-Dialekt)*. Berlin: Wissenschaft & Technik 1998.

Simoni, Zef: *Fjalor gjermanisht-shqip*. Tiranë: 8 Nëntori 1978.

Smiltniece, Gunta: "Lietvārds (substantīvs)". In: Auziņa, Ilze et al.: *Latviešu valodas gramatika*. Rīga: LU Latviešu valodas institūts 2015, pp. 324–369.

Stassen, Leon: "Predicative possession". In: Haspelmath, Martin et al. (eds.): *The World Atlas of Language Structures*. Oxford: Oxford University Press 2005, pp. 474–477.

Stassen, Leon: *Predicative Possession* (= Oxford Studies in Typology and Linguistic Theory). Oxford: Oxford University Press 2009.

Stolz, Thomas et al.: *Split Possession. An Areal-Linguistic Study of the Alienability Correlation and Related Phenomena in the Languages of Europe* (= Studies in Language Companion Series 101). Amsterdam, Philadelphia: Benjamins 2008.

Stolz, Thomas/Levkovych, Nataliya: "'haben – besitzen – gehören' – Einige Vorüberlegungen zu einer typologisch inspirierten Komplettierung der Possessionsforschung". In: Urdze, Aina/Levkovych, Nataliya (eds.): *Linguistik im Nordwesten. Beiträge zum 8. Nordwestdeutschen Linguistischen Kolloquium, Bremen, 13.-14. November 2015* (= Diversitas Linguarum 42). Bochum: Brockmeyer 2017, pp. 95–127.

Stolz, Thomas & Levkovych, Nataliya: "I don't belong! Preliminary thoughts on the typology of belong-constructions". In: Johanson, Lars/Mazzitelli, Lidia/Nevskaja, Irina (eds.): *Linguistic Possession. New Insights from the Language of Europe and North and Central Asia*. Amsterdam, Philadelphia: Benjamins, to appear (accepted).

Tabačenko, Ljudmila: *Pristavočnye pozicionnye glagoly v istorii russkogo jazyka: poliaspektnyj analiz*. Avtoreferat 2011.

Thomas, R.J./Bevan, Gareth A./Donovan, Patrick J.: *Geiriadur Prifysgol Cymru. Cyfrol III: M-Rhywyr*. Caerdydd: Gwasg Prifysgol Cymru 2007.

Tompa, József: *Kleine ungarische Grammatik*. Leipzig: Enzyklopädie 1972.

Ultan, Russell: "Towards a typology of substantival possession". In: Greenberg, Joseph H. (ed.): *Universals of Human Language. Volume 4: Syntax*. Stanford: Stanford University Press 1978, pp. 11–50.

Vinogradov, V.V.: *Istorija slov: okolo 1500 slov i vyraženij i bolee 5000 slov, s nimi svjazannych* (otvetstvennyj redaktor N. Ju. Švedova). Moskva: Tolk 1994.

Webster's: *Webster's Encyclopedic Unabridged Dictionary of the English Language*. New York, etc.: Random House 1994.

Winkler, Eberhard: *Udmurtische Grammatik* (= Veröffentlichungen der Societas Uralo-Altaica 83). Wiesbaden: Harrassowitz 2011.

Wood, Francis A.: Notes on Latin Etymologies. In: *Classical Philology* 7 (3), 1912, pp. 302–334.

Luka Szucsich

Burgenland Croatian as a Contact Language

Abstract: This contribution discusses Burgenland-Croatian, spoken in Austria and in some villages in Hungary and Slovakia, as a contact language focusing especially on its contact with German in its Bavarian variety. Besides giving an overview of grammatical properties of Burgenland-Croatian, which have only recently become subject to language change due to general bilingualism, with German being the dominant language, this paper investigates three morpho-syntactic phenomena which can be attributed to long-standing language contact or, more generally speaking, to areal convergence with German. However, for all these phenomena, alternative explanations of their origin are more or less plausible. This shows that it is methodologically very challenging to decide on the exact source of the morpho-syntactic linguistic patterns in Burgenland-Croatian, which can be viewed as exceptional in the context of South Slavic languages.

Keywords: language contact, Burgenland-Croatian, German, morpho-syntactic change, bilingualism

1 Introduction

Burgenland Croatian (B-Cr) can be viewed as predestined to explore language contact phenomena. This can be attributed to it being a language spoken by a minority population surrounded by speakers mainly of German, but also of Hungarian and to a lesser degree Slovak and Romani. This has indeed been the perspective in Slavic linguistics for certain linguistic phenomena in B-Cr. With hardly any exceptions (e.g., Hadrovics 1958 for particle verbs which will be discussed below in section 4.2), language contact phenomena going beyond lexical borrowings have been only briefly mentioned without being investigated in more detail. This especially holds for morpho-syntactic phenomena and aspects of word order of certain grammatical elements, e.g., clitics, verb-placement, or question formation.

So, in her very short chapter on the syntax of a particular B-Cr subdialect[1], Koschat (1978, pp. 131–132) only marginally mentions that certain uses of PPs and of the infinitive can be explained by assuming "German influence". Other instances of syntactic phenomena attributed by her to language contact with

---

1   Koschat (1978) focuses on the dialect of the so-called *Poljanci* (cf. section 2.2 for details), especially on the dialect spoken in the village of Pajngrt/Baumgarten.

German more represent lexical borrowings, e.g., the use of *jedan*[2] 'one' instead of *neki* 'some' as an indefinite marker, or the use of *ča za* 'what for' instead of *kakov* 'which' as an exclamative marker. On good empirical grounds, Browne (2010) rejects an analysis of exceptional clitic order patterns in B-Cr by resorting to language contact with German or Hungarian. He also briefly mentions the tendency for verb second (V2) placement in B-Cr and states that language contact with German is likely to be at play, although V2 is not grammaticalized in B-Cr (Browne 2010, p. 35).

Before I turn to linguistic phenomena likely attributable to language contact in section 4, I will give a brief historical overview of the origins of the B-Cr population in Austria, Hungary and Slovakia (historically also in Czechia) in section 2. In section 3, I will discuss some linguistic characteristics of B-Cr which only recently became subject to change especially in younger speakers of the community. The phenomena discussed in section 4 seem to be more systematic and can be attested in different older sources, e.g. in spoken data collected by Koschat (1978) or Neweklowsky (1978) or even in data collected during the Wenker survey carried out in Austria at the end of the 1920s, beginning of the 1930s (cf. Fleischer 2017, Szucsich to appear).

2 Historical Overview: The Burgenland Croats

The B-Cr population is concentrated in the Austrian federal state of Burgenland (B-Cr *Gradišće*). The endonym for Burgenland Croats is traditionally *Hrvati* 'Croats' and *hrvatski* 'Croatian' for the language. Outside the academic context, the term *Gradišćanski Hrvati* 'Burgenland Croats' and *gradišćanskohrvatski* 'Burgenland Croatian' is mostly used by the members of the community in contexts where a distinction has to be made relative to Croats/Croatian from Ex-Yugoslavia. The term Burgenland Croats also extends to Croatian speaking populations in other parts of Austria, Hungary, Slovakia and (up to the end of the Second World War) Southern Moravia (Czech Republic) historically connected to the Croats in Burgenland.

At present, B-Cr is spoken in the Austrian federal state of Burgenland, in Vienna by speakers originating from Burgenland, in a narrow strip of Western

---

2   Contrary to Koschat (1978) and Neweklowsky (1978) or other authors who use a phonetic transcription of B-Cr examples, I will use a normalized orthography, since I am not concerned with phonetic/phonological phenomena in this paper. The orthography used in this paper is oriented on the B-Cr standard and deviates from this only in cases where written sources are cited.

Hungary along the Austrian border, and in some villages near the Slovakian capital Bratislava. Historically, B-Cr was much more widespread and included speakers in many villages in Lower Austria along the river Danube between Vienna and Bratislava, in numerous villages along the river Morava on both sides of the Austrian-Slovakian border, in some villages North of Vienna near Stockerau, in some villages in Southern Moravia near Mikulov, in some villages in the Western Carpathian mountains North-East of Bratislava, and in a few more villages in today's Western Hungary, e.g. near Mosonmagyaróvár (for a detailed description of B-Cr historical settlements cf. the seminal work by Breu 1970, and also Koschat 1978 und Neweklowsky 1978, Tobler 1983a, 1983b, 1986, Österreichische Rektorenkonferenz 1989).

2.1 Migration History of the Burgenland Croats

The B-Cr population can be traced back to a settlement of Croats in the 16th century which, contrary to popular belief, was largely not characterized by unorganized population movement or even flight from the Ottomans who were advancing in South Eastern Europe and threatening the kingdom of Hungary. The settlement was more due to recruitment by the lords of manors in the Western part of the kingdom of Hungary, which included today's Austrian federal state of Burgenland as well as Slovakia, and in Eastern parts of Lower Austria and Southern Moravia which was by then a Habsburg possession, cf. Breu (1970), Tobler (1983a, 1983b, 1986, 1992). At the beginning of the 16th century a lot of villages in the abovementioned areas, especially those dominated by wheat production, were deserted or partly deserted. A sometimes contested assumption is that the reason for the widespread depopulation was foremost economic, not least driven by a fall in the price for wheat in the 14th and 15th century, cf. Abel (1978) for this claim extending to all of Central Europe.

The so-called agricultural crisis[3] of the late middle ages in Central Europe is often identified as the main cause of an overall decline in population and a rural exodus (cf. Abel 1978). Additional factors as to why the border areas between

---

3 Some scholars question the mono-causal explanations in Abel (1978), especially the assumed central role of the plague pandemics in the 14th and 15th centuries (Dolle 1994), and the fact that Abel (1978) does not take into account the regional differences in the agricultural economy of Central Europe (cf. Rösener 1992). Despite criticism, there is consensus that certain areas of Central Europe were stricken by heavy depopulation. In the case in question, it is remarkable that wheat producing villages in particular suffered the most (cf. Tobler 1983a, 1983b, 1986).

Hungary and Austria were especially affected by depopulation are probably the military campaigns of the so-called Black Army of Matthias Corvinus (Hunyadi Mátyás) at the end of the 15th century and the raids preceding and accompanying the attempt to conquer Vienna by the Ottomans at the beginning of the 16th century. The lords of manors were interested in resettling the depopulated villages, which led to the recruitment of Croatian settlers from areas which were under threat of being occupied by the advancing Ottoman forces (see below for the region of origin of the Burgenland Croats). The settlers were granted tax exemptions for a couple of years and could bring their movable property to their new homes free of duty. In some regions—especially in Lower Austria—privileges were rather quickly revoked (for the historical background of the settlements cf. among others Breu 1970, Tobler 1983a, 1983b, 1986, 1992, Österreichische Rektorenkonferenz 1989).

Some of the villages settled by Croats were only partly Croatian-speaking from the beginning of the settlement, cf. Breu (1970)[4]. This is true for most of the villages east of Vienna along the banks of the river Danube. Most of the villages in this area had received Croatian settlers, but only few were predominantly or entirely Croatian-speaking in the 16th century (cf. Brabec 1966, Breu 1970, Tobler 1983a, 1983b). A similar situation holds for the settlements along the banks of the river March/Morava which today forms, for the most part, the border between Austria and Slovakia (cf. Breu 1970). Many of the villages in Lower Austria had already assimilated by the 18th century, some by the 19th century. There were only few speakers of B-Cr left in the 20th century. The last reports of elderly Croatian speakers can be found for the 1950s (cf. Brabec 1966 on a 74 year old speaker in Loimersdorf/Limištrof). But in quite a few cases, during the first decades and centuries after the settlement of the Croats, a German or Hungarian minority in a village would linguistically assimilate to the Croatian majority (cf. Breu 1970 for details). Furthermore, the villages settled by Croats never formed a contiguous and coherent language area. To the contrary, B-Cr was, and still is, spoken in language pockets in the immediate vicinity of German, Hungarian and Slovak speaking villages. As a consequence, language contact was and is ubiquitous.

---

4   Tobler (1992) gives a detailed account of the relations between German speaking and Croatian speaking inhabitants in the mixed village of Kittsee/Gjeca in the district of Neusiedl am See/Niuzalj in Burgenland. Despite the fact that Kittsee/Gjeca was a linguistically mixed village for centuries—between the wars one third of the inhabitants still declared themselves to be Croats—, heavy assimilation to German started only after the Second World War.

The areas the B-Cr settlers originate from are situated to the South of the contemporary Croatian capital Zagreb in the border region of today's Croatia and Bosnia between the rivers Kupa and Una and South of the river Una. Roughly speaking, these areas lie within the triangle formed by the cities Karlovac (Croatia), Kutina (Croatia) and Bihać (Bosnia), cf. esp. Neweklowsky (1978, 2010) and also Breu (1970), Koschat (1978), Österreichische Rektorenkonferenz (1989), and Houtzagers (2008).

At that time, those parts of Croatia were predominantly Čakavian speaking, bordering Kajkavian and Štokavian dialects. Čakavian, Kajkavian and Štokavian are the three main dialect groups of what is called Bosnian/Croatian/Serbian (BCS) in the academic discourse, i.e. former Serbo-Croatian. All these dialects have numerous sub-dialects. Čakavian dialects differ significantly from the dialects which form the basis for Standard BCS which are so-called Neo-Štokavian dialects, i.e., Štokavian dialects with a great amount of linguistic innovations. Today, in Croatia, Čakavian dialects are mainly spoken in Istria, on the Adriatic islands and in a very thin strip along most parts of the Croatian mainland coastal line (for further details concerning the dialectal situation in the BCS area, especially in Croatian, cf. among many others, Ivić 1956, Lisac 2003, 2009, and Lončarić 1996). B-Cr shares many linguistic features with Čakavian dialects in Croatia, cf. section 3 below.

2.2 The Present Situation of Burgenland Croatian

At present, there are an estimated 20–25.000 B-Cr speakers in Burgenland and, additionally, a further 15.000 estimated speakers in Vienna, Hungary and Slovakia. There are only a few speakers left in the Czech Republic after their forced resettlement from three villages in Southern Moravia to different parts in then-Czechoslovakia, mainly to Northern Moravia (1948–1950), cf. Vašková (2013). Currently, assimilation is accelerating in most B-Cr language pockets, although over the past few decades the institutional situation of B-Cr much improved in different social spheres like (i) administration (confirmation as a recognized official language in Burgenland by the constitutional court in 1987), (ii) education (foundation of a bilingual secondary school/gymnasium in Oberwart/Borta in 1992), and (iii) media (increased radio and television broadcasting in B-Cr in state owned media, ORF, since the 1980s), cf. for details different contributions in Holzer/Münz (1993), Szucsich (2000). Furthermore, since the late 1970s, a B-Cr standard language of its own based on the dominant Čakavian Ikavian-Ekavian dialects started to be developed, cf. Weilguni (1984), Holzer/Münz (1993), Szucsich (2000), which ended in the creation of dictionaries and

reference grammars, cf. among others Bencsics et al. (1982, 1991) and Sučić (2003). Still, linguistic assimilation is progressing. In comparison to the present figures given above, after the First World War, there were an estimated 60.000 B-Cr speakers in Burgenland and, additionally, an estimated 30.000 speakers in Vienna, Lower Austria, Hungary, and Czechoslovakia.

3 Linguistic Characteristics of Burgenland Croatian

3.1 Dialectal Division of Burgenland-Croatian

The vast majority of Burgenland Croats speak Čakavian dialects, and among the Čakavians so-called Ikavian-Ekavian dialects prevail. The latter term refers to phonological reflexes of the so-called Old Slavic *Jat*, a phoneme most probably pronounced as a near open unrounded front vowel [æ] or as a diphthong [ɪe]. In the Slavicist literature, this phoneme is commonly represented as /ě/. In Ikavian-Ekavian dialects, the former *Jat*-vowel is realized either as an /i/ or as an /e/ mostly depending on the consonant following the vowel (cf. Koschat 1978, Neweklowsky 1978, 2010), e.g., *ml**i**ko* 'milk' vs. *leto* 'year, summer' This picture is further complicated by the fact that in almost all B-Cr dialects the long mid vowels /eː/ and /oː/—and in some subdialects also stressed short mid vowels— were subject to diphthongization, i.e., /ˈl� ieto/ for *leto* 'year, summer' or /mo̯oj/ for *moj* 'my'. At present, the Čakavian, Ikavian-Ekavian dialects are mainly spoken by the so-called *Haci* and *Poljanci* in Northern Burgenland and by the *Dolinci* in Central Burgenland, as well as by speakers in Western Hungary and Slovakia. These dialects exhibit moderate Kajkavian and—to a lesser degree—Štokavian dialectal features. Historically, Croats in Lower Austria and Southern Moravia also spoke these dominant dialects, cf. Koschat (1978), Neweklowsky (1978, 2010), Houtzagers (2008).

In Southern Burgenland, one portion of the Croats—the so-called Southern Čakavians—also speaks Čakavian dialects, but Ikavian ones, i.e., the Old Slavic *Jat* is represented by /i/ irrespective of its position (*ml**i**ko*, *l**i**to*). The second portion—the so-called *Štoji* and *Vlahi*—speaks Štokavian dialects which, however, exhibit several Čakavian dialectal features. Except for two villages with Ikavian-Ekavian reflexes of the *Jat* (see above), Štokavians speak Ikavian dialects. Ekavian reflexes can be found even in Ikavian villages, although only with certain lexemes, e.g., with the lexeme *seno* /ˈs̯ieno/ 'hay', cf. Neweklowsky (2010, p. 164). The transition between Čakavian and Štokavian dialects is often gradual due to the significant amount of Čakavian features in Štokavian dialects,

cf. Koschat (1978), Neweklowsky (1978, 2010), and Houtzagers (2008) for further details. There are also several villages with Štokavian dialects in Hungary bordering Southern Burgenland. Besides these larger dialectal groups, there are two Kajkavian villages south of lake Neusiedl on the Hungarian side of the border and one Kajkavian village in Slovakia. These dialects also exhibit certain Čakavian features, cf. Koschat (1978), Neweklowsky (1978, 2010), Houtzagers (2008).

3.2 Morphological and Syntactic Properties of Burgenland-Croatian

Before turning to morpho-syntactic phenomena in B-Cr which are most likely induced by language contact or—more broadly speaking—by widespread bilingualism among speakers of B-Cr, I want to briefly discuss some morphological and syntactic characteristics of B-Cr which are at least partly similar to those of other Slavic languages. Nevertheless, some of the properties are subject to ongoing changes especially in younger speakers. The data which I will present in order to exemplify non-canonical patterns in recent B-Cr are not systematically elicited, though. The examples showing deviation from expected linguistic behavior mostly stem from authentic production of B-Cr speakers in television and radio broadcasting.

3.2.1 Case in Burgenland Croatian

B-Cr has a differentiated case system with six cases which resemble the system in many Slavic languages and which can be said to have been inherited from common Slavic: *nominative*, *accusative*, *genitive*, *dative*, *instrumental*, and *locative*. The often-so-called non-syntactic *vocative* is missing in spoken B-Cr with only some remnants left, e.g., *Bože* 'Oh God!'. There are some unusual patterns with dative NPs, accusative NPs and locative NPs which may appear without prepositions in a directional (dative, accusative) and locative (locative) meaning, cf. (1) for examples with bare accusative and bare locative NPs with a directional and a locative meaning, respectively.

(1) a.   Tat-a   je   iša-o   crikv-u.
          Dad$_{\text{N.SG}}$ AUX$_{\text{3.SG}}$ go$_{\text{LPT-m.SG}}$ church$_{\text{A.SG}}$
          'Dad went to church.'
     b.   Tat-a   je   crikv-i.
          Dad$_{\text{N.SG}}$ COP$_{\text{3.SG}}$ church$_{\text{L.SG}}$
          'Dad is in church.'

Thus, despite language contact to languages with case systems different from the Slavic one[5], B-Cr did not undergo reorganization of its case system or the inventory of case markers, e.g., by the levelling of markers. Quite to the contrary, case markers in B-Cr are more differentiated than in (innovative) neo-Štokavian BCS making B-Cr more "archaic" in this respect, cf. Tab. 1 for two declension patterns for the lexemes *grad* 'city, town; castle' and *žena* 'woman; wife' (cf. Sučić 2003 for a full description of Standard B-Cr morphology, and Težak/Babić 1992 for a full description of BCS).

B-Cr has the same number of declension classes in the plural as in the singular whereas BCS generalizes declension patterns in the plural at least with respect to so-called non-structural cases. BCS has the ending *ima/ama* for INST, DAT, LOC in all declension classes in the plural. In B-Cr, the 1st declension class has the ending *i* for the INST and the 2nd declension class has the ending *ami*. Similar contrasts appear in the LOC (*i* vs. *ā*) and in the DAT (*om* vs. *ām*). As already mentioned, BCS has a strong tendency for syncretism in the plural paradigm. Partly, this holds also for the markers in the singular within one paradigm. Strictly speaking, there is no distinction between DAT and LOC in the singular in any of the BCS declension classes, so these two cases actually collapse into one case. In contrast, within and across paradigms, B-Cr does not have the same amount of syncretic forms as BCS. In many declension classes, there are different markers for DAT, INST and LOC in the plural. In the 2nd declension class, there are three different markers: *ām*, *āmi*, *ā*. In some declension classes in the singular, some subdialects have a distinct marker for DAT and LOC. This situation in B-Cr is very similar to that in Slovene or Czech and some other Slavic languages.

Of course, B-Cr also exhibits syncretisms in the plural paradigms. This is true for LOC.PL and INST.PL of the 1st declension class (*grad-i*).[6] This syncretism is due to the overall loss of the consonant /h/ in the coda of the last syllable of a word

---

5   German has a four-case system (*nominative*, *accusative*, *genitive* and *dative*) with markers more often appearing on the determiner, than on the noun itself (of course, with exceptions). Besides cases resembling the Slavic system, Hungarian has an extensive array of different locational markers appearing on the noun (*inessive*, *adessive*, *illative*, *allative*, to name just a few).

6   The marker for NOM.PL is also *i*. Despite the identical endings, there is still a phonological distinction between NOM.PL and INST.PL/LOC.PL forms: In most cases, the stressed stem vowel bears a different tone, cf. *grȁd-i* vs. *grád-i* in Tab. 1.

form. The ending of the LOC.PL developed from an *ih* into an *i*. The same holds for the ending of the LOC.PL of the 2nd declension class: *āh > ā* which still differs from the ending of the NOM.SG consisting of a short vowel (*a*).[7]

Tab. 1: Two declension patterns in B-Cr compared to BCS (cf. Sučić 2003, pp. 92/116, Težak/Babić 1992, pp. 86/88/91)[a]

	B-Cr		BCS		B-Cr		BCS	
	Sg	Pl	Sg	Pl[b]	Sg	Pl	Sg	Pl
Nom	grâd	grâd-i	grâd	grȁd-ov-i	žèn-a	žèn-e	žèn-a	žèn-e
Gen	grâd-a	grâd-ov	grâd-a	grȁd-ōv-ā	žèn-ē	žên	žèn-e	žén-ā
Dat	grâd-u	grâd-om	grâd-u	grȁd-ov-ima	žèn-i	žèn-ām	žèn-i	žèn-ama
Acc	grâd	grâd-e	grâd	grȁd-ov-e	žèn-u	žèn-e	žèn-u	žèn-e
Loc	grâd-u	grád-i	grád-u	grȁd-ov-ima	žèn-i	žèn-ā	žèn-i	žèn-ama
Inst	grâd-om grád-i		grâd-om grȁd-ov-ima		žèn-ōm	žen-ámi	žèn-om	žèn-ama
	1st declension (*o*-stems)				2nd declension (*a*-stems)			

[a]Tab. 1 gives just two declension patterns out of many more in B-Cr. I left out the vocative which is non-existent in spoken B-Cr. Its existence in the prescriptive grammar Sučić (2003) is due to the fact that the authors of the standardization moderately oriented themselves toward the neo-Štokavian standard (cf. Weilguni 1984, Szucsich 2000). I also simplified the patterns by giving only one variant in those cases where two slightly different forms are mentioned in Sučić (2003) for some of the cells in the paradigm (e.g., *grȁd-i* in the INST.PL instead of *grád-i*). Furthermore, there is, of course, morphological variation in different subdialects of B-Cr, e.g., in some Čakavian Ikavian-Ekavian subdialects, the ending for the INST.SG in the 2nd declension class (*a*-stems) given in table 1 is usually *u* not *om*. In most Čakavian Ikavian-Ekavian subdialects, the ending for the LOC.SG in the 1st declension class (*o*-stems) is typically *i* not *u*, cf. Neweklowsky (1978) who gives a detailed description of dialectal morphological variation in B-Cr.
Some forms also differ with respect to the length of vowels in the stem and/or in the ending (long or short) or with respect to tone (rising and falling). The latter distinction is confined to stressed vowels, the former extends to unstressed vowels. Unstressed vowels may be either short or long. In BCS, they are more often short, in particular B-Cr subdialects, pretonic vowels may be long (cf. Neweklowsky 1978). Standard BCS makes a four-way distinction with stressed vowels: Vowels may be short falling (ȁ), short rising (à), long falling (â), and long rising (á). Standard B-Cr has a three-way distinction (cf. Neweklowsky 1978, Sučić 2003). Only long vowels make an intonational distinction (â vs. á). The short vowel is represented as 'ȁ'.
[b]B-Cr does not have the stem extension *ov/ev* with monosyllabic and some bisyllabic masculine nouns systematically found in BCS.

---

7   The loss of the /h/ in the absolute coda also led to syncretisms in the adjectival declension. In the B-Cr standard, the ending of the GEN.PL is written with *ih*. The /h/ is never pronounced in spoken B-Cr, though, which makes it syncretic to the ending of the NOM.PL of masculine adjectives.

A rather recent development esp. in young speakers of B-Cr is the loss or attrition, and in some cases the re-organization of case markers. These changes probably correlate with developments caused by the loss of distinctions in the phonological system, especially the distinction with respect to vowel tones and, to a certain degree, the quantitative distinction in unstressed syllables. These developments increase the number of syncretisms within and across paradigms, e.g., the collapse of NOM.PL with LOC.PL and INST.PL in the 1st declension class, cf. Tab. 1 and footnote 8. Thus, the case system which was stable for a few centuries lost its stability in the last decades, at least for some speakers.

A second explanation for the re-organization of the case system in B-Cr might be general tendencies for grammatical systems in language contact situations, especially in bilingual communities with dominant majority languages (L2). However, newly established (emerging) grammatical patterns need not be due to simple structural transfer from dominant languages. Polinsky (2006) shows that, with less proficient speakers of so-called American Russian[8], case markers of marked cases disappear compared to baseline Russian (= monolingual Russian spoken in Russia). In many American Russian speakers, dative markers are replaced by accusative markers in contexts of so-called GOAL-marking, e.g., with indirect objects of di-transitive verbs, and the accusative collapses with the nominative establishing what Polinsky (2006) calls UNMARKED CASE. Except for fossilized forms, less proficient speakers of American Russian lose the instrumental and the prepositional (locative) entirely, and in some instances the genitive as well, cf. Polinsky (2006) for details.

For most B-Cr speakers in Austria, German became the dominant language. This holds also for many speakers between the ages of 50–60. In a lot of cases, it is even difficult to speak of B-Cr as the only L1 (again, not only for the youngest generation). In the last few decades, German has been acquired simultaneously with B-Cr by most bilinguals and has become more dominant with the onset of formal education in school. The same holds for Hungarian regarding B-Cr speakers in Hungary. Similar to American Russian, this may lead to the replacement of marked cases by less or entirely unmarked cases, cf. (2).

---

8   American Russian is spoken by first or second-generation immigrants to the U.S. exposed to Russian as an L1 in earlier years who left a Russian speaking environment and switched to English as the primary language at the age of 3–11 (cf. Polinsky 2006).

(2) To   je   zbog   #ver-u.[9]
    This$_{N.SG}$ COP$_{3.SG}$ because-of   faith$_{A.SG}$
    'This is because of the faith.'
    (adult speaker from Parndorf/Pandrof; source: ORF, Dobar dan Hrvati,
    27.08.2017)

The authentic example in (2) was produced by an adult speaker from a village
(Parndorf/Pandrof) where the process of assimilation began rather early in the
20th century. In this case, the genitive marker required by the preposition *zbog*
'because of' is replaced by the accusative marker: *ver-u*$_{ACC}$ instead of *ver-e*$_{GEN}$
'faith'. In all case hierarchies proposed for Slavic languages, the genitive ranks
higher, i.e. it is more marked, than the accusative, cf. Jakobson (1936) or Müller
(2004) among many others for case hierarchies for Russian.

Unfortunately, there is no systematic survey available on possible changes in
the linguistic behavior of B-Cr speakers with regard to case markers. Phenomena
as in (2) have to be systematically investigated in order to make substantial claims
about non-canonical case patterns in the more recent forms of B-Cr. According
to cursory observation, less proficient speakers indeed do not use cases ranked
particularly high on cross-linguistic case hierarchies, like the INST and the LOC,
e.g., by avoiding prepositions governing these cases.

3.2.2 Clitics in Burgenland Croatian

Similar to BCS, B-Cr has a full-fledged system of (future and perfect tense) auxil-
iary, pronominal and reflexive clitics, cf. Browne (2010) for a detailed description.
In declarative clauses, clitics canonically appear in the so-called second position,
although marginally clitics may also occupy the clause-initial position.[10] The
term 'second position' for B-Cr does not cover the same range of possibilities as
in BCS, though. In BCS, clitics may appear after the first prosodic word within
a constituent. In contrast, in B-Cr declarative clauses, clitics have to follow the
whole constituent (XP) if in second position—which is the canonical position,
as already mentioned, cf. (3). The ungrammatical position in (3)—indicated by

---

9   From a prescriptive view, the marker *u* is ungrammatical in this context, which is usu-
    ally indicated by an asterisk: *. Since this might be a case of a recurrent, non-canonical
    pattern, the annotation # is chosen to indicate a deviant marker.
10 Neweklowsky (1978, p. 253) provides some examples with clause-initial clitics from
    authentic spoken texts. However, those instances are far less frequent than non-initial
    placement of clitics. Some of the examples even sound deviant, if presented in isolation.

the asterisk within the brackets—is grammatical in BCS. In B-Cr, the perfect tense auxiliary *je* has to follow the entire NP *naša sestra* 'our sister'—indicated by the asterisk before the brackets.[11] Furthermore, in an example as in (3), a clause-initial position of the perfect tense auxiliary would sound very awkward, if not entirely ungrammatical. So, there are obviously restrictions concerning the initial position of clitics in declarative clauses.

(3) Naš-a   (*je)   sestr-a   *(je)   doš-l-a.
     Our$_{f.N.SG}$ AUX$_{3.SG}$ sister$_{N.SG}$ AUX$_{3.SG}$ come$_{LPT-f.SG}$
     'Our sister arrived.'

Contrary to the overall ban of clause-initial clitics in BCS, B-Cr polar questions containing perfect and future tense verbs have virtually exclusively clause-initial auxiliary clitics, cf. (4). This resembles the situation in Slovene polar questions, cf. Franks/King (2000, p. 40), except for the additional option in Slovene to have an initial question particle *ali* followed by an auxiliary clitic, which is absent in B-Cr root questions. In B-Cr, only embedded polar questions are introduced by the related question particle/complementizer *ali*. Crucially, polar questions in Čakavian dialects in Croatia behave similarly and allow for initial clitics in root clauses. Additionally, these dialects more readily than B-Cr allow for clause-initial clitics in declarative clauses, cf. Kalsbeek (1998), Lisac (2009).

(4) Je   ur   doš-l-a?
     AUX$_{3.SG}$ already   come$_{LPT-f.SG}$
     'Has she come yet?'

Thus, with certain peculiarities with respect to the ordering of clitics within the clitic cluster, B-Cr clitics do not behave differently from other Slavic clitic systems, especially Western South Slavic ones.

---

11 For further details concerning restrictions with respect to the clitic cluster see Browne (2010) who investigates written B-Cr texts. The main ordering principles within the clitic cluster are: future tense auxiliary clitics (*ću* 'will$_{1.SG}$'; *ćeš* 'will$_{2.SG}$'; *će* 'will$_{3.SG}$'; *ćemo* 'will$_{1.PL}$'; *ćete* 'will$_{2.PL}$'; *ćedu* 'will$_{3.PL}$') and perfect tense auxiliary clitics except for the 3.sg clitic *je* 'be$_{3.SG}$', viz. *sam* 'be$_{1.SG}$'; *si* 'be$_{2.SG}$'; *smo* 'be$_{1.PL}$'; *ste* 'be$_{2.PL}$'; *su* 'be$_{3.PL}$' precede pronominal clitics like *mu* 'him$_{DAT}$'; *ga* 'him$_{ACC}$'; *joj* 'her$_{DAT}$'; *ju* 'her$_{ACC}$'; etc. For a more detailed description of ordering restrictions within the clitic cluster see Browne (2010), but also Neweklowsky (1978) who investigates spoken texts. In contrast to BCS, there are no non-clitic auxiliary forms, i.e., there are no B-Cr counterparts to BCS non-clitic auxiliaries like *jesam* 'be$_{1.SG}$' (vis-à-vis clitic *sam*).

Similar to the developments concerning the case system, only recently one can observe non-canonical placement of clitics esp. in younger speakers. The example in (5a) was produced by a child aged 6 growing up in Vienna, which makes it plausible that German is the dominant language. In this example, the auxiliary clitic *smo* and the accusative reflexive clitic *se* are split, i.e. they do not form a cluster. More importantly, the reflexive clitic appears clause-finally following the second non-clitic constituent *jigrali*.

(5)   a.   #... a   onda   smo   jigra-l-i   se.
           ... and   then   AUX$_{1.PL}$ play$_{LPT-m.PL}$ REFL$_{ACC}$
           '... and then we were playing.'
           (speaker, age 6, from Vienna/Beč; source: ORF, Dobar dan Hrvati, 23.07.2017)

      b.   ... a   onda   smo   se   jigra-l-i.
           ... and   then   AUX$_{1.PL}$ REFL   play$_{LPT-m.PL}$
           '... and then we were playing.'

Again, I chose the diacritic # over the asterisk (*) to indicate the deviance of the example, because this kind of clitic placement, especially with the reflexive clitics *se*$_{ACC}$ and *si*$_{DAT}$, was recurrent in the speech of this child in this particular TV-cast. Proficient adult speakers judge examples like (5a) to be ungrammatical, (5b) with a clitic cluster in second position being the grammatical variant for proficient speakers. As with changes concerning the case system, there is no systematic survey on clitic placement with less proficient, especially younger speakers of B-Cr.

3.2.3   Verbal Aspect in Burgenland Croatian

As Koschat (1978, p. 121) and Neweklowsky (1978, p. 244) observed for the speakers recorded in the 1960s and 1970s, B-Cr retained the grammaticalized aspectual distinction between perfective (pfv.) and imperfective (ipfv.) verb forms typical for Slavic languages. A lot of verbs form what is called in the Slavicist literature aspectual pairs, which are morphologically distinct. Exceptions to verbal lexemes containing morphologically distinct aspectual pairs are biaspectual (e.g., *hasnit(i)*$_{PFV/IPFV}$[12] 'to be of use, to help'), stative (e.g., *znat(i)*$_{IPFV}$ 'to know')

---

12   The infinitive is given as *t(i)*, because with certain exceptions (cf. Neweklowsky 1978), the infinitive in spoken B-Cr dropped the final vowel *i*, i.e. the ending is *t*, not *ti*. In

and atelic verbs denoting a process (e.g., *spat(i)*$_{IPFV}$ 'to sleep')[13]. The latter two classes are *imperfective tantum*. The morphological means for forming aspectual pairs correspond to those of other Slavic languages, especially those in BCS, and resemble derivational strategies rather than inflectional (cf. Lehmann 2003 for the morphology of aspect in Russian).

The first morphological strategy is secondary imperfectivization which can be achieved either by adding an imperfectivizing suffix *-(i)v-* to a pfv. base as in (6a) or by changing the so-called thematic vowel as in (6b). Thematic vowels are added to the stem and determine conjugation patterns. Thus, the change from *i* to in (6b) also always changes the conjugation class the verb form belongs to. Especially the latter strategy is often accompanied either by the insertion of a stem vowel or by a qualitative change of the stem vowel, cf. (6c) for a stem vowel insertion. Furthermore, secondary imperfectivization is often accompanied by consonant alternations in the coda of the root as in (6d). Equivalent consonant alternations are well attested in other Slavic languages. The second productive strategy is prefixation of ipfv. bases which leads to perfectivization as in (6e).[14] Depending on lexical information of the verbal base, different prefixes may serve as perfectivizers.

(6) a. *z-li-t(i)* 'spill$_{PFV}$' – *z-li-v-a-t(i)* 'spill$_{IPFV}$'
     b. *hit-i-t(i)* 'throw$_{PFV}$' – *hit-a-t(i)* 'throw$_{IPFV}$'
     c. *ot-pr-i-t(i)* 'open$_{PFV}$' – *ot-pir-a-t(i)* 'open$_{IPFV}$'
     d. *ot-pust-i-t(i)* 'release$_{PFV}$' – *ot-pušć-a-t(i)* 'release$_{IPFV}$'
     e. *pis-a-t* 'write$_{IPFV}$' – *na-pis-a-t* 'write$_{PFV}$'

As with the phenomena discussed in the previous two sections, with some B-Cr speakers—not exclusively younger ones—the use of aspectual forms is deviant from that of proficient speakers, at least in specific contexts. The distinction between pfv. and ipfv. aspect is still rather stable in cases in which the pfv. verb denotes a semelfactive event and the ipfv. counterpart is canonically interpreted

---

the grammars for standardized B-Cr as well as in printed media, the infinitive always has the ending *ti* (cf. Sučić 2003 among others).

13 The latter class can, of course, serve as the morphological base for so-called aktionsart-derivations like the phasal aktionsart-verb *za-spati* 'to fall asleep' which are perfective. Aktionsart-derivations are not considered to represent pfv. partners of ipfv. verbs, cf. Isačenko (1968), Schwall (1999).

14 In the context of the present paper, it is not essential to discuss the controversial question, whether prefixation can form purely aspectual pairs, cf. Isačenko (1968), Schwall (1999) among many others.

as an iterated event (*skoč-i-t(i)* 'jump$_{PFV}$ (once)' vs. *skak-a-t(i)* 'jump$_{IPFV}$ (repeatedly)'). The same is true for most accomplishment verbs, i.e. telic predicates which denote situations involving a dynamic preparatory phase and a clearly definable result state which is applied to the internal argument in its totality (e.g., the direct object with transitive verbs) as in (6e).[15] With accomplishments, the aspectual contrast is especially clear in the past tense. In episodic readings, the ipfv. forms canonically denote ongoing processes, where the result state is not yet reached (progressive reading), whereas pfv. forms canonically mark that the result state is part of the asserted situation.

The situation is different, though, with verbs of motion. Some speakers use the pfv. verb forms in episodic present tense contexts, where proficient speakers would only use the ipfv. form, cf. (7).

(7) (On) dojd-e.   (*instead of*:   id-e/gr-e)
     he   come$_{PFV.PRS-3.SG}$ come$_{IPFV.PRS-3.SG}$
     'He is coming.'

Cases as in (7) and similar usage of pfv. forms of verbs of motion can be observed rather frequently. With even less proficient speakers, the deviant usage especially of pfv. verb forms in contexts where one would expect ipfv. forms is extended to other verb classes.

4 Burgenland Croatian and Language Contact

As already mentioned in section 2, B-Cr is in especially close contact to German and also to Hungarian since the migration of its speakers to their new homeland in the 16th century. Contact-induced linguistic phenomena have been discussed in a great number of works on B-Cr (cf. Koschat 1978, Neweklowsky 1978 among others), but little attention has been paid to (morpho-)syntactic phenomena (but cf. Hadrovics 1958, Tornow 1993). In the following, I will discuss selected (morpho-)syntactic phenomena in order to illustrate systematic linguistic traits which are an integral part of the grammar of proficient speakers. Hence, these phenomena differ from recent non-canonical patterns in less proficient speakers of B-Cr cursorily discussed in the previous section.

---

15 The dynamic preparatory phase in this case is the process of writing something. The result state is the object's property of being totally written, cf. Krifka (1989), Szucsich (2005) among many others.

4.1 Doubly Filled COMP

Similar to Standard English and Standard German, but in contrast to Scandinavian languages, such as Swedish, and also to certain varieties of German (Bavarian) and English (Belfast English), most Slavic languages do not exhibit what has been often dubbed 'doubly filled COMP' (but cf. Bacskai-Atkari 2018 for cases in Czech and Slovene discussed below). The so-called filter on doubly filled COMP amounts to a ban of structures which contain a *wh*-phrase in a left peripheral position (like *which dish* in example (8)), and a complementizer like *that*, cf. the ungrammatical example in Standard English in (8). In the context of B-Cr, it is especially important to mention that Standard BCS does not allow for a doubly filled COMP, at least not with the C-element *da* 'that', cf. (9).

(8)   *I wonder *which dish that* they picked.

(9)   *Ne   zna-m,   *koga   da*   je   traži-o.
        NEG   know$_{PRS-1.SG}$ who$_A$ that   AUX$_{3.SG}$ look-for$_{LPT-m.SG}$
        'I do not know, who he was looking for.'

Interestingly, B-Cr at least marginally allows for the violation of the doubly filled COMP filter as can be seen in (10).

(10)   ?Joži   ne   zna,   *koga   da*   je   vidi-o.
        Joži   NEG   know$_{PRS-3.SG}$ who$_A$ that   AUX$_{3.SG}$ see$_{LPT-m.SG}$
        'Joži does not know who he saw.'

If one superficially considers the most frequent linear order in Slavic constituent questions (*wh*-questions), they behave similarly to those in English or in German: The constituent marked with the interrogative morphology (*wh*-phrase) canonically appears at the left edge of the clause. However, it is difficult to tell for root questions, whether *wh*-phrases indeed occupy a position within the C-domain (CP or another phrase in the very left periphery of the clause) or lower in the clausal structure (within the TP/IP). Meyer (2004) and Dyakonova (2009), among others, show that *wh*-phrases in root clauses in many Slavic languages may appear in non-initial positions. However, in embedded questions, the fronting of *wh*-phrases like *koga* 'who$_A$' in B-Cr to the C-domain is claimed to be obligatory in languages like BCS, cf. Bošković (2002). The same seems to be true in B-Cr. Furthermore, the element *da* 'that' is arguably a complementizer category occupying C^0, since it canonically introduces declarative embedded clauses in B-Cr. Hence, examples like (10) indeed represent instances of doubly filled COMP.

It is a well known fact that Southern German varieties, e.g. Bavarian dialects spoken in most parts of Austria, also violate the doubly filled COMP filter. A relevant example is given in (11). This may suggest that the fact that B-Cr allows for doubly filled COMP is due to transfer from local German varieties which exhibit similar phenomena.

(11) I hob koa Ahnung, mid wos fia-ra Farb dass-a zfrien waar. [Bav]
     I have no idea   with what for-a color that-he content would-be
     'I have no idea with what color he would be happy.'

                                                (Bayer/Brandner 2008, p. 88)

On the other hand, it also has been attested that there are violations of the doubly filled COMP filter in certain Slavic languages which are partly geographically close to the areas of origin of the Burgenland Croats. Bacskai-Atkari (2018) reports for Slovene and Czech that both options for constituent questions exist: one with a fronted *wh*-phrase accompanied by a C⁰-element *da* (Slv) or *že* (Cz) immediately following the *wh*-phrase, and one without the respective C⁰-element. Interestingly, not only embedded, but also root *wh*-questions in Slovene and Czech may appear with the complementizer, compare the pairs of root and embedded questions in (12) and (13) for Slovene.

(12) a. *Kdo*   prid-e?   [Slv]
      Who$_N$ come$_{PRS:3.SG}$
      'Who is coming?'
   b. Vpraša-l   je,   *kdo*   prid-e.
      asked$_{LPT-m.SG}$ AUX$_{3.SG}$ who$_N$ come$_{PRS:3.SG}$
      'He asked who was coming.'

(13) a. *Kdo*   *da*   prid-e?
      Who$_N$ that   come$_{PRS:3.SG}$
      'Who is said to be coming?'
   b. ?Vpraša-l je,   *kdo*   *da*   prid-e.
      asked$_{LPT-m.SG}$ AUX$_{3.SG}$ who$_N$ that come$_{PRS:3.SG}$
      'He asked who was said to be coming.'

                                          (Bacskai-Atkari 2018, pp. 11–12)

However, as Bacskai-Atkari (2018) reports, the versions with and without the complementizer *da* are not fully equivalent doublets. As the translation of the examples in (12) and (13) indicates, the insertion of the complementizer *da* results in an interpretive difference from ordinary *wh*-questions and renders what is often called an echo question interpretation. So, the sentence in (13a) is

an appropriate reaction to a statement such as *"Peter is coming"* if the addressee of the statement did not properly understand it. The sentence in (13b) is the embedded version of an echo question. Its markedness is due to the fact that it is relatively difficult to find contexts in which an embedded echo question is felicitous.

The same interpretive difference has been reported for Czech, cf. Gruet-Skrabalova (2011) and Kašpar (2015), where the complementizer *že* 'that' can be inserted in root questions as in (15a) as well as in embedded questions as in (15b). Again, the respective questions are interpreted as echo-questions.

(14) a. *Kdo*   přije-l?   [Cz]
         Who$_N$ arrive$_{LPT-m.SG}$
         'Who arrived?'
      b. Pta-l-a   se,   *kdo*   přije-l.
         asked$_{LPT-f.SG}$ REFL   who$_N$ arrive$_{LPT-m.SG}$
         'She asked who arrived.'

(15) a. *Kdo*   *že*   přije-l?
         Who$_N$ that   arrive$_{LPT-m.SG}$
         'Who is said to have arrived?'
      b. ˀPta-l-a   se,   *kdo*   *že*   přije-l.
         asked$_{LPT-f.SG}$ REFL   who$_N$ that   arrive$_{LPT-m.SG}$
         'She asked who was said to have arrived.'

                                                    (Bacskai-Atkari 2018, p. 8)

Bavarian does not show any of these interpretive differences with sentences containing a doubly filled COMP. It also does not readily allow for root questions with an overtly realized C⁰-element *dass* 'that'. A sentence as in (16a) is only very marginally acceptable in hearsay readings which suggests that a matrix clause was elided. A V2 *wh*-question as in (16b) is entirely ungrammatical.

(16) a. ˀˀMid wos fia-ra Farb   dass-a   zfrien   waar?   [Bav]
          with what for-a   color   that-he   content would-be
          'With what color would he be happy?'
      b. * Mid wos fia-ra Farb   (waar)   dass-a   waar   zfrien?
          with what for-a   color   would-be   that-he   would-be content
          'With what color would he be happy?'

Interestingly, B-Cr rather patterns with Bavarian in all these respects. As Bavarian, B-Cr does not exhibit an interpretational effect of an inserted COMP

category. As the translation in (10) indicates, the respective sentence has a run-of-the-mill interpretation of an embedded question. Furthermore and in sharp contrast to Slovene and Czech, B-Cr does not allow for violations of the doubly filled COMP filter in root questions, cf. (17a). Since doubly filled COMP constructions are extremely marginal with a simple subject *wh*-phrase in embedded clauses as in (17b), one could assume that the ungrammaticality of (17a) follows from independent restrictions excluding only a subject *wh*-phrase from being followed by a complementizer. However (17c) shows that object *wh*-phrase in root clauses also defy the insertion of the complementizer *da*, although the clause mirrors the grammatical embedded clause in (10).

(17) a. **Ki   da   je   doša-o?*
         Who$_N$ that   AUX$_{3.SG}$ arrive$_{LPT\text{-}m.SG}$
         'Who (is said to have) arrived?'

     b. ??*Pita-l-a   je,   *ki   da*   je   doša-o.
         Asked$_{LPT\text{-}f.SG}$ AUX$_{3.SG}$ who$_N$ that   AUX$_{3.SG}$ arrive$_{LPT\text{-}m.SG}$
         'She asked who (was said to have) arrived.'

     c. **Koga   da   je   vidi-o?*
         Who$_A$ that   AUX$_{3.SG}$ see$_{LPT\text{-}m.SG}$
         'Who (is said) did he see?'

There is another interesting finding reported by Bayer/Brandner (2008). Doubly filled COMP constructions in Bavarian and Allemanic dialects of German get significantly better, if the *wh*-element is phrasal, i.e. not a word size element, as in (11). Bayer/Brandner (2008) present results from a systematic survey showing that simplex *wh*-element may not be followed by a complementizer, cf. (18).

(18) *I woass aa ned, *wer dass* allas am Sunndoch in da Kiach   gwen is. [Bav]
     I know too not   who that   all   at Sunday   in the church been is
     'I don't know either who all has been to church on Sunday.'

                                                              (Bayer/Brandner 2008, p. 89)

B-Cr exhibits similar improvement of doubly filled COMP constructions with phrasal *wh*-elements. In contrast to cases like (10) which are felt to be slightly deviant, complex *wh*-elements like the accusative *wh*-phrase *ku knjigu* 'which book' preceding the complementizer *da* are perfectly grammatical, cf. (19a). Still, if the very same *wh*-phrase appears in a root question, the sentence becomes ungrammatical, cf. (19b).

(19) a. Joži ne zna, *k-u* *knjig-u* *da* je kupi-o.
    Joži NEG know$_{PRS-3.SG}$ which$_{A.SG}$ book$_{A.SG}$ that AUX$_{3.SG}$ buy$_{LPT-m.SG}$
    'Joži doesn't know which book he bought.'

   b. **Ku* *žen-u* *da* je vidi-o?
    Which$_{A.SG}$ woman$_{A.SG}$ that AUX$_{3.SG}$ see$_{LPT-m.SG}$
    'Which woman (is said) did he see?'

One has to point out, though, that there are certain differences between B-Cr and Bavarian. In contrast to B-Cr, Bavarian along with Allemanic does not readily tolerate simplex, word-like *wh*-elements like accusative *wen* 'who(m)' preceding the complementizer *dass*, i.e., there is no subject object asymmetry as in B-Cr. Despite this, B-Cr and Bavarian share the common properties of not allowing for doubly filled COMP constructions with root questions and of not showing any interpretive effects of doubly filled COMP constructions (i.e., the respective sentences allow for a reading as a "simple" embedded question). These data strongly suggest that the B-Cr type of doubly filled COMP constructions are rather the product of language contact to Bavarian varieties of German spoken in all parts of Eastern Austria than the result of a common development with other Slavic languages which are geographically relatively close to B-Cr.

4.2 "Prefixal Composition" (Phrasal Verbs)

In the B-Cr morpho-syntactic system, certain adverbs and also some prepositions which have no prefix status in other Slavic languages, especially in BCS, are re-analyzed as verbal prefixes. Contrary to ordinary, Slavic-type prefixed verbs which also exist in B-Cr[16] these "new" prefixed verbs behave more like phrasal verbs in German, especially with respect to their word order patterns, cf. below. In German linguistics they are called *trennbare Verben* 'separable verbs'. Since, for a Slavic language, phrasal verbs are extremely exceptional, an explanation involving language contact is very plausible.

Phrasal verbs appear in B-Cr no later than the 18th century. Hadrovics (1958) discusses examples from B-Cr texts from the 18th and the 19th century claiming that this phenomenon was due to previous language contact with Hungarian in the areas of origin of the Burgenland Croats (cf. section 2.2). However, the

---

16 As already discussed in section 3.2.3, B-Cr employs prefixation to form pfv. verbs from simplex ipfv. ones: *del-a-t(i)*$_{IPFV}$ – *u-del-a-t(i)*$_{PFV}$ 'to make, to work'. There are also multiple prefixations, although not as frequent as in other Slavic languages: *po-za-pir-a-t(i)*$_{PFV}$ 'to close one after another'. These prefixes behave like cognate prefixes in other Slavic languages.

examples he gives are extremely rare and marginal in the Croatian texts of that time period. Besides, the Burgenland Croats' original settlement area was not in immediate contact with Hungarian speaking areas. As will be shown below, at least the current linguistic behavior of phrasal verbs in B-Cr points towards linguistic transfer from German.

According to Hadrovics' (1958) findings, originally, this morpho-syntactic innovation was semantically restricted to concrete spatio-temporal meanings. But during the 20th century this derivational strategy spread to more abstract meanings, and now additional adverbs, sometimes direct loans from German, are employed to derive phrasal verbs. This development clearly can be analyzed as an integration of morphological patterns from German, cf. (20) which obviously is a calque from German with an originally spatio-temporal phrasal prefix *najpr* 'forward' used to derive a complex verb with a non-spatio-temporal meaning.

(20)   *si najpr zet(i)* 'to resolve, to take up' (cf. Ger. *sich vornehmen*)

(21) comprises a list of very frequent B-Cr phrasal prefixes with German (and Hungarian) equivalents, cf. also Hadrovics (1958) and, additionally, Tornow (1993), Szucsich (2000).

(21)   a.   *doli* 'down'   —   nieder/unter/hinunter/herunter (Hu. *le, alá*)
        b.   *gori* 'up'   —   auf/hinauf/herauf (Hu. *fel*)
        c.   *kraj* 'away'   —   weg (Hu. *elfélre*)
        d.   *najpr* 'forward'   —   vor/hervor (Hu. *elő*)
        e.   *najzad* 'back'   —   zurück (Hu. *vissza*)
        f.   *nutar* 'in(to)'   —   ein/hinein/herein (Hu. *be, bele*)
        g.   *prik* 'over'   —   über/hinüber/herüber (Hu. *át*)
        h.   *skupa* 'together'   —   zusammen (Hu. *össze*)
        i.   *van* 'out'   —   aus/hinaus/heraus (Hu. *ki*)

More evidence for the claim that the behavior of phrasal verbs in B-Cr is due to transfer from German stems from the peculiarities of word order with these verbs. B-Cr in general allows for verb final structures in contexts of VP-focus or wide (sentential) focus even with sentences containing transitive verbs and full subject and object NPs. In these contexts, an SVO order is also available, although an SOV order is slightly preferred, cf. (22) with both orders in a context inducing wide focus.

(22)   Context: *Tome is crying. A asks: What happened? B answers:*

a. Tome   je   knjig-u   zgubi-o.
   T.$_N$ AUX$_{3.SG}$ book$_{A.SG}$ lose$_{LPT-m.SG}$
   'Tome lost the book.'
b. Tome je zgubio knjigu.

With phrasal verbs, the right peripheral verb placement becomes even more natural if not the only available option under VP- or wide focus, cf. (23a) and (24a). The verb without the particle/phrasal prefix may be placed in front of the object NP, cf. (23b) and (24b), although this word order is more restricted and far less natural under VP- or wide focus than in cases like (22b) with non-phrasal verbs. The intermediate position of the entire phrasal verb between the subject (plus clitics) and the object is deviant, at least in a neutral context, cf. (23c) and (24c). The appropriate context for this word order is one triggering contrastive focus on the object *kameke* 'stones' and *nož* 'knife'. The particle/phrasal prefix on its own is completely excluded from the intermediate position, cf. (23d) and (24d) which is surprising given the often so-called "free word order" which is otherwise attested in B-Cr, too. The data in (23d) and (24d) shows that phrasal verbs in B-Cr cannot be syntactically split, if the phrasal prefix precedes the verb. Similar to German, the prefix can only be left behind, if the verb is moved to the left, preferably in the right-peripheral position.

(23) a. Jive   je   kamek-e   *doli*   hita-o.
          J.$_N$ AUX$_{3.SG}$ stone$_{A.PL}$ down   throw$_{LPT-m.SG}$
          'Jive was throwing down stones.'
      b. Jive je hitao kameke *doli*.
      c. $^?$ Jive je *doli* hitao kameke.
      d. * Jive je *doli* kameke hitao.

(24) a. Katica je   nož   *kraj*   vrg-l-a.
          K.$_N$ AUX$_{3.SG}$ knife$_{A.SG}$ away   put$_{LPT-f.SG}$
          'Katica put away the knife.'
      b. Katica je vrgla nož *kraj*.
      c. $^?$ Katica je *kraj* vrgla nož.
      d. *Katica je *kraj* nož vrgla.

If the verb appears in the left-peripheral position within a root clause, e.g. in *pro*-drop contexts, the phrasal prefix has to remain in the clause-final position as in (25a). This is true, if there is no contrastive focus available, i.e., for a sentence answering the question *What did Jive do?* The preverbal position as in (25b) and

the position preceding the object as in (25c) are deviant, though not entirely ungrammatical, since they are marginally allowed in marked contexts. A sentence as in (25c) requires a context inducing contrastive focus on the object NP accompanied by a corresponding contrastive intonation on *kameke*. Examples like (25b) also require a special intonation often described as a bridge contour which is information structurally associated with contrastiveness as well (cf. among others Mehlhorn 2004 for bridge contours in Russian).

(25) a.   Hitao   je   kamek-e   *doli*.
         Throw$_{LPT-m.SG}$ AUX$_{3.SG}$ stone$_{A.PL}$ down
         'He was throwing down stones.'

     b. $^{??}$*Doli* hitao je kameke.
     c. $^{??}$Hitao je *doli* kameke.

Interestingly, phrasal prefixes in B-Cr do not entirely fit into the Slavic morphological patterns for aspectual derivations. There is only a weak tendency that they combine with simplex, unprefixed verbs. Furthermore, the phrasal prefixes behave inconsistently with respect to perfectivization. Depending on the lexical root, they may combine with simplex, imperfective verbs with unclear results concerning the aspectual status of the derivation.[17] This may lead to more or less synonymous doublets with Slavic-type prefixed verbs, cf. (26a, b) where the phrasal prefix cannot be combined with the prefixed base. The prefixed verb is in some cases already less frequent in use compared to the phrasal verb derivation, e.g. *spustit(i)* compared to *doli pustit(i)* in (26b). With some verbal roots, phrasal prefixes also may attach to already prefixed perfective verbs and to unprefixed perfective verbs (cf. Szucsich 2000 for further details). In these cases, a derivation with the simplex base would be ungrammatical, cf. (26c, d). In cases like (26c, d), the phrasal prefixes *doli* and *kraj* do not really contribute an additional meaning component, i.e. the derivations might be called pleonastic.

(26) a.   zletit(i)   > $^{??}$gori zletit(i)   >   gori letit(i)   'to fly up'
     b.   spustit(i)   >   *doli spustit(i)   >   doli pustit(i)   'to lower'

---

17   Tornow (1993) claims that they tend to replace regular prefixes. Pairs like *doli hit-i-t(i)*$_{PFV}$ – *doli hit-a-t(i)*$_{IPFV}$ 'to throw down' derived from *hit-i-t(i)*$_{PFV}$ – *hit-a-t(i)*$_{IPFV}$ 'to throw' speak against this interpretation.

c.	spast(i)	>	doli spast(i)	>	*doli past(i)	'to fall off'
d.	odsić(i)	>	kraj odsić(i)	>	*kraj sić(i)	'to chop off'

In sum, the behavior of phrasal verbs, which is unseen in other Slavic languages, strongly suggests that this phenomenon is contact-induced. Besides, the word order regularities make German the most plausible source of linguistic transfer in this case, viz. (i) the strong tendency of the phrasal verb to the right periphery (verb last), and (ii) the strong tendency to "separation" of the phrasal prefix, if the verb is fronted.

4.3 Purposive (Final) Infinitival Clauses

In this section, another phenomenon will be discussed which is often classified as a product of language contact with German: the frequent use of infinitival purposive/final clauses in B-Cr, sometimes called consecutive in the literature on German, in contexts where BCS and other Slavic languages more frequentlyuse certain types of finite clauses (often containing non-factual morphology) or event nominalizations, cf. Koschat (1978, p. 131), Neweklowsky (1978, p. 46) for some examples of infinitival purposive clauses in B-Cr. In most Slavic standard languages, clauses containing an infinitival form of the verb are not used in contexts of adverbial adjunct clauses in general, which also holds true for Standard BCS. In contrast, German purposive clauses are often introduced by the element *um* and an infinitive which is accompanied by the particle *zu*.

The B-Cr material discussed in this section stems from the first half of the 20th century (so-called "Wenker material") in order to show that the patterns in question are not developments attributable to very recent language change. During the subsequent data collection for the German Linguistic Atlas (*Deutscher Sprachatlas*) in Austria at the end of the 1920s and in the early 1930s, the newly created federal state of Burgenland was also included in the dialectological survey. As with other multilingual survey areas, the material includes data from non-German speaking sites among them are also 10 more or less complete questionnaires containing B-Cr linguistic data, cf. Szucsich (to appear) for details on the B-Cr questionnaires and on linguistic phenomena found therein.[18]

---

18 There are also two Hungarian questionnaires from Burgenland, a questionnaire in Yiddish from Burgenland, and Slovene questionnaires from 13 sites in Carinthia.

The questionnaires do not cover all the dialectal sub-groups of B-Cr[19] (cf. section 3.1 for the dialectal division of B-Cr), although for the phenomenon of purposive clauses, this does not play an important role.

The core of the questionnaire form for the subsequent dialectological survey in Austria consisted of the so-called 40 Wenker sentences which were, by and large, already used in the original dialectological surveys from 1876–1887 in the German Empire, cf. Fleischer (2017). The relevant sentence in the questionnaire containing a purposive clause is sentence 16 in Wenker's survey, given in (27) with an English translation. The markers introducing the purposive clause (*um*) and accompanying the infinitive (*zu*) are in italics. The latter is positioned between the phrasal prefix *aus-* and the verbal root *trink-* which is the only possible position of this marker in the infinitive of phrasal verbs in German.

Wenker sentence 16:

(27)   Du bist noch nicht groß genug, *um* eine Flasche Wein aus*zu*trinken,
Du mußt erst noch ein Ende/etwas wachsen und größer werden.
'You aren't big enough to drink a whole bottle of wine. You have to grow some more first and get bigger'.

The expected pattern for the equivalent of the purposive clause … *um eine Flasche Wein auszutrinken* … in a Slavic language would be a clause introduced by a complementizer and a subjunctive element/morpheme. Subjunctive markers, in turn, require the verb of the purposive clause to appear as a so-called *l*-participle which otherwise forms the most widespread past tense in Slavic languages (in North Slavic languages, virtually the only past tense form). The subjunctive marker evolved from the aorist of the copula verb *be* which has the form *by-* or *bi-*, depending on the Slavic language, and which may be accompanied by agreement morphology. The latter is the case in BCS and Polish, but not in Russian and B-Cr which employ an indeclinable particle *by* and *bi* respectively.

In B-Cr questionnaires, a pattern for Wenker sentence 16 is rather frequent which seems to be adopted from German. Out of the 10 questionnaires, the relevant purposive clause is realized (i) 3 times with a preposition-like element *za* 'for' and an infinitive, (ii) 2 times with a bare infinitive without any preposition-like element or a complementizer, (iii) once with the element *za* without a verb

---

19   Some of the smaller sub-groups are not represented at all. This is true for the Southern Čakavians (district of Güssing) or the so-called *Haci* (district of Neusiedl) which are Northern Ikavian-Ekavian Čakavians. However, the latter are linguistically very close to the *Poljanci* (district of Eisenstadt and Mattersburg) which are represented in two questionnaires.

(*za flošu vina* 'for a bottle of wine'), (iv) two times with the complementizer *da* 'that' marked with the subjunctive particle *bi* and an *l*-participle (the expected form), and (v) two times with the complementizer *da* without the subjunctive marker and with a finite verb form (in questionnaire 42676 Zillingtal a pfv. present tense form *spiješ* 'drink₂.SG', in questionnaire 43863 Podgoria a present tense form of the modal verb *moreš* 'can₂.SG' and the infinitive of the main verb *popit* 'drink PFV.'). In sum, the infinitival strategy, (i) and (ii), is slightly more numerous than the finite one, (iv) and (v).

The examples in (28) are representative of the two strategies. The first one, (28a), is from questionnaire 42681 Trausdorf and represents the expected finite embedding with a subjunctive marker on the complementizer, the second one, (28b), is from questionnaire 42740 Kroatisch Minihof representing the infinitival strategy.[20]

(28) a. Još ni-si   dost   velik, da-bi floš-u vin-a spi-l.
       yet NEG-AUX₂.SG enough big m.SG that SBJV bottle A.SG wine G.SG drink LPT-m.SG
       'You aren't big enough to drink a bottle of wine.'
    b. Još ni-si   dost   velik za   floš-u vin-a spi-t.
       yet NEG-AUX₂.SG enough big m.SG for   bottle A.SG wine G.SG drink INF
       'You aren't big enough to drink a bottle of wine.'

The surveys for the *Deutscher Sprachatlas* also covered other Slavic speaking areas, among which were: Polish speaking areas in the original surveys in the German Empire and areas where Carinthian dialects of Slovene are spoken in the Austrian ones. The data from the respective questionnaires shows that the B-Cr strategy to express purposive subordinate relations (a preposition/complementizer and an infinitive of the embedded verb) resembling the German one are indeed exceptional. To illustrate the expected "Slavic" pattern, one Polish (53478 Rombark) and two Slovene examples from Carinthia (44317 Zell-Pfarre and 44154 Thörl-Maglern) are given in (29) which contain a non-factual, subjunctive marker (*-by-* and *b(i)*) and the historical *l*-participle (marked with *-ł-* in Polish and *-l-* in Standard Slovene, *-v-* in the two questionnaires).

---

20 In the B-Cr examples in (28), as well as in the Polish and Slovene examples in (29), the original orthography from the questionnaires is maintained and does not in all instances correspond to the respective standardized orthographical rules.

(29) a. [...] abyś   flaszk-ę   win-a   wypi-ł.   [Po]
      COMP-SBJV$_{2.SG}$ bottle$_{A.SG}$ wine$_{G.SG}$ drink$_{LPT-m.SG}$
      '[...] to drink a bottle of wine.'

   b. [...] da   b   ano   flašk-o   vin-a   pupiv.   [Slv]
      that   SBJV   one$_{A}$ bottle$_{A.SG}$ wine$_{G.SG}$ drink$_{LPT-m.SG}$
      '[...] to drink a bottle of wine.'

   c. [...] da   bi   na   flaschk-o   Vin-a   papiv.
      that   SBJV   one$_{A}$ bottle$_{A.SG}$ wine$_{G.SG}$ drink$_{LPT-m.SG}$
      '[...] to drink a bottle of wine.'

Thus, language contact with German is a plausible option to account for the exceptional status of the B-Cr data in (28). However, one has to be cautious with a final verdict on the source of this morpho-syntactic phenomenon. Certain Čakavian dialects in Istria—in this case Central Istrian dialects of the Northwest Čakavian dialect group—display the same patterns in purposive clauses, cf. (30).[21]

(30)   [...]   (kosir)   ča   rab-i   za   čisti-t [...]. [IstrČak]
           (bent-knife$_{N.SG}$)   what   be-needed$_{PRS.3.SG}$ for   clean$_{INF}$
       '[...] a bent knife which is needed to clean [...].'   (Kalsbeek 1998, p. 290)

Since infinitival purposive clauses are limited to certain Čakavian dialects and otherwise unknown in the BCS area, it seems plausible that they are a product of language contact with Italian. Italian forms purposive clauses with the help of the prepositions *per* 'for, to' or *a* 'to, at' and the infinitive. Linguistic transfer from (Venetian) Italian to Čakavian dialects and other South Slavic varieties spoken along the Adriatic coast is well attested. Kalsbeek (1998) states that the dialect she describes in her monography was subject to influence by Venetian Italian and other Romance varieties for centuries. This influence resulted not only in lexical borrowings which are abundant in coastal dialects of Croatian, but also in phonological and syntactic transfer. B-Cr has a larger number of lexical borrowings from Italian (e.g., *facol* 'handkerchief', it. *fazzoletto*; *durat(i)* 'to last', it. *durare*; cf. Koschat 1978 for further examples), although far less than coastal dialects of Croatian. But otherwise, phonological or morpho-syntactic patterns

---

21   The diacritics for accentual characteristics of stressed vowels in the example in (30) and (31) given in Kalsbeek (1998) are omitted, since they are not relevant for our purpose.

resembling Italian ones (and differing from Slavic patterns) are not attested in the literature on B-Cr.

Besides, infinitives in purposive clauses in B-Cr are much more limited than in the said Istrian Čakavian dialects. In Istrian Čakavian, infinitival embedded clauses introduced by *za* are also used to express so-called rationale clauses, a subgroup of purposive clauses, which differ from proper purpose clauses in various ways (cf. Bach 1982 for details). A relevant example is given in (31).

(31) Ja ć-u   ga   otpača-t za ga   da-t   t-e   ž^{j}ensk-e^{22}.
    [IstrČak]

   I $FUT_{1.SG}$ $CL_{m.A.SG}$ $empty_{INF}$ for $CL_{m.A.SG}$ $give_{INF}$ $that_{f.D.SG}$ $woman_{D.SG}$
   'I will empty it (the pan), to give it to that woman.' (Kalsbeek 1998, p. 290)

In B-Cr, an equivalent of (31) could not be used with an infinitival embedded clause. The most natural way would be with a finite clause containing a subjunctive marker ([…] *da-b(i) ga da-o t-oj žen-i.* '[…] $that_{SBJV}$ $CL_{m.A.SG}$ $give_{LPT-m.SG}$ $that_{f.D.SG}$ $woman_{D.SG}$.'), i.e. with the expected "Slavic-type" strategy, or by completely breaking up the subordinate syntactic relation. So, if B-Cr indeed brought along the infinitival syntactic pattern for purposive clauses from its area of origin, one had to assume that it lost subparts of the pattern in the new homeland.

5 Summary

B-Cr shows several grammatical phenomena in morpho-syntax which justifiably can be called exceptional in the context of more or less common Slavic or South Slavic grammatical properties. Some of those phenomena, viz. (i) doubly filled COMP, (ii) phrasal verbs and their syntactic behavior, and (iii) a subgroup of purposive clauses, which are all already present in B-Cr for a longer period of time, can be best analyzed as induced by language contact. However, the present contribution aimed at showing that in most cases it is very difficult to ultimately decide on the exact contact language. Nevertheless, for all the discussed cases German, especially in its Bavarian variety, seems to be the most plausible candidate.

---

22 The ending for the DAT.SG of nouns of the 2nd declension class in *-a* (mostly feminine nouns) and of the feminine adjectives sharply differs from that in B-Cr (*t-oj žen(sk)-i* 'that $woman_{D.SG}$') which equates the ending in Standard BCS.

References

Abel, Wilhelm: *Agrarkrisen und Agrarkonjunktur. Eine Geschichte der Land- und Ernährungswirtschaft Mitteleuropas seit dem hohen Mittelalter*. Parey: Hamburg. 1978.

Bach, Emmon: "Purpose clauses and control". In: Jacobson, Pauline/Pullum, Geoffrey K. (eds.): *The Nature of Syntactic Representation*. Reidel: Dordrecht. 1982, pp. 35–57.

Bacskai-Atkari, Julia: "Doubly filled COMP in Czech and Slovenian interrogatives". In: Lenertová, Denisa/Meyer, Roland/Šimík, Radek/Szucsich, Luka (eds.): *Advances in formal Slavic linguistics 2016*. Language Science Press: Berlin. 2018, pp. 1–23.

Bayer, Josef/Brandner, Ellen: "On wh-head-movement and the doubly-filled-comp filter". In: Chang, Charles B./Haynie, Hannah J. (eds.): *Proceedings of the 26th West Coast Conference on Formal Linguistics*. Cascadilla Proceedings Project: Somerville. 2008, 87–95.

Bencsics, Nikolaus et al.: *Deutsch-burgenländischkroatisches-kroatisches Wörterbuch/Nimško-gradišćanskohrvatsko-hrvatski rječnik*. Ed. Roetzer: Eisenstadt–Zagreb. 1982.

Bencsics, Nikolaus et al.: *Gradišćanskohrvatsko-hrvatsko-nimški rječnik/Burgenländischkroatisch-kroatisch-deutsches Wörterbuch*. GZH: Zagreb–Eisenstadt. 1991.

Bošković, Željko: "On multiple wh-fronting". *Linguistic Inquiry* 33(3), 2002, pp. 351–383.

Brabec, Ivan: "Govor podunavskih Hrvata u Austriji". *Hrvatski dijalektološki zbornik* 2, 1966, pp. 29–118.

Breu, Josef: *Die Kroatensiedlung im Burgenland und den anschließenden Gebieten*. Deutike: Wien. 1970.

Browne, Wayles: "Syntactic studies in Burgenland Croatian: the order of clitics". *Balkanistica* 23, 2010, pp. 21–41.

Dolle, Josef. "Zu der Theorie einer "spätmittelalterlichen Agrarkrise". Eine kritische Untersuchung am Beispiel des Altkreises Göttingen". *Göttinger Jahrbuch* 42, 1994, pp. 55–94.

Dyakonova, Marina: *A phase-based approach to Russian free word order*. LOT Publications: Utrecht. 2009.

Fleischer, Jürg: *Geschichte, Anlage und Durchführung der Fragebogen-Erhebungen von Georg Wenkers 40 Sätzen: Dokumentation, Entdeckungen und Neubewertungen*. Georg Olms: Hildesheim/Zürich/New York. 2017.

Franks, Steven/King, Tracy Holloway: *A handbook of Slavic Clitics*. Oxford University Press: New York. 2000.

Gruet-Skrabalova, Hana: "Czech questions with two wh-words". In: Kosta, Peter/Schürcks, Lilia (eds.): *Formalization of grammar in Slavic languages*. Peter Lang: Frankfurt a.M et al. 2011, pp. 179–192.

Hadrovics, Lászlo: "Adverbien als Verbalpräfixe in der Schriftsprache der burgenländischen Kroaten". *Studia Slavica Academiae Scientiarum Hungaricae* 4(1–2), 1958, pp. 211–249.

Holzer, Werner/Münz, Rainer (eds.): *Trendwende? Sprache und Ethnizität im Burgenland*. Passagen: Wien. 1993.

Houtzagers, Peter: "On Burgenland Croatian isoglosses". In: Houtzagers, Peter/Kalsbeek, Janneke/Schaeken, Jos (eds.) *Dutch Contributions to the Fourteenth International Congress of Slavists, Ohrid: Linguistics (SSGL 34)*. Rodopi: Amsterdam–New York. 2008, pp. 293–331.

Isačenko, Aleksandr V.: *Die russische Sprache der Gegenwart. Formenlehre*. Niemeyer: Halle. 1968.

Ivić, Pavle: *Dijalektologija srpskohrvatskog jezika: Uvod u štokavsko narečje*. Matica srpska: Novi Sad. 1956.

Jakobson, Roman: "Beitrag zur allgemeinen Kasuslehre". Reprint in: *Form und Sinn. Sprachwissenschaftliche Betrachtungen*. Fink: München. (1936) 1974, pp. 77–124.

Kalsbeek, Janneke: *The Čakavian Dialect of Orbanići near Žminj in Istria*. Rodopi: Amsterdam. 1998.

Kašpar, Jiří: "On pied-piping and feature percolation". In: Zybatov, Gerhild et al. (eds.): *Slavic grammar from a formal perspective*. Peter Lang: Frankfurt a.M. et al. 2015, pp. 245–258.

Koschat, Helene: *Die čakavische Mundart von Baumgarten im Burgenland*. Österreichische Akademie der Wissenschaften: Wien. 1978.

Krifka, Manfred: *Nominalreferenz und Zeitkonstitution: Zur Semantik von Massentermen, Pluraltermen und Aspektklassen*. Fink: München. 1989.

Lehmann, Volkmar: "Grammatische Derivation (Aspekt, Genus verbi, Komparation, Partizip und andere Phänomene zwischen Flexion und Wortbildung) ". In: Berger, Tilman/Gutschmidt, Karl (eds.): *Funktionale Beschreibung slavischer Sprachen. Beiträge zum XIII. Internationalen Slavistenkongress in Ljubljana*. Sagner: München, 2003, pp. 139–162

Lisac, Josip: *Hrvatska dijalektologija 1*. Golden marketing–Tehnička knjiga: Zagreb. 2003.

Lisac, Josip: *Hrvatska dijalektologija 2*. Golden marketing–Tehnička knjiga: Zagreb. 2009.

Lončarić, Mijo: *Kajkavsko narječje*. Školska knjiga: Zagreb. 1996.

Mehlhorn, Grit: "The Prosodic Pattern of Contrastive Accent". In: Steube, Anita (ed.): *Information Structure: Theoretical and Empirical Aspects*. de Gruyter: Berlin–New York. 2004, pp. 241–258.

Meyer, Roland: *Zur Syntax der Ergänzungsfrage. Empirische Untersuchungen am Russischen, Polnischen und Tschechischen*. Sagner: München. 2004.

Müller, Gereon: "A distributed morphology approach to syncretism in Russian noun inflection". In: Arnaudova, Olga et al. (eds.): *Formal Approaches to Slavic Linguistics: The Ottawa Meeting 2003*. Michigan Slavic Publications: Ann Arbor. 2004, pp. 353–373.

Neweklowsky, Gerhard: *Die kroatischen Dialekte des Burgenlandes und der angrenzenden Gebiete*. Österreichische Akademie der Wissenschaften: Wien. 1978.

Neweklowsky, Gerhard: *Die Sprache der Burgenländer Kroaten. Jezik gradišćanskih Hrvatov*. ZIGH: Trajštof/Trausdorf. 2010.

Österreichische Rektorenkonferenz (ed.): *Bericht der Arbeitsgruppe* "*Lage und Perspektive der Volksgruppen in Österreich*". Böhlau: Wien. 1989.

Polinsky, Maria: "Incomplete Acquisition: American Russian". *Journal of Slavic Linguistics* 14(2), 2006, pp. 191–262.

Rösener, Werner: *Agrarwirtschaft, Agrarverfassung und ländliche Gesellschaft im Mittelalter*. Oldenbourg: München. 1992

Schwall, Ulrike: *Aspektualität. Eine semantisch-funktionelle Kategorie*. Gunter Narr: Tübingen. 1999.

Sučić, Ivo (ed.): *Gramatika gradišćanskohrvatskoga jezika*. ZIGH: Željezno/Eisenstadt. 2003.

Szucsich, Luka: "Das Burgenlandkroatische: Sprachwandel, Sprachverfall, Sprachverschiebung und Sprachassimilation". In: Zybatow, Lew (ed.): *Sprachwandel in der Slavia. Die slavischen Sprachen an der Schwelle zum 21. Jahrhundert*. Peter Lang: Frankfurt a.M. et al. 2000, pp. 853–875.

Szucsich, Luka: "Kompositionalität von Aspekt und ihre Bedingungen: Ein russisch-finnischer Vergleich". In: Wingender, Monika (ed.): *Linguistische Beiträge zur Slavistik XII*. Sagner: München, 2005, pp. 171–195.

Szucsich, Luka: "Die burgenlandkroatischen Wenkerbögen". In: Jürg Fleischer et al. (ed.): *Minderheitensprachen und Sprachminderheiten. Deutsch und seine Kontaktsprachen in der Dokumentation der Wenker-Materialien*. Georg Olms: Hildesheim/Zürich/New York. To appear.

Težak, Stjepko/Babić, Stjepan: *Gramatika hrvatskog jezika*. Školska knjiga: Zagreb. 1992.

Tobler, Felix: *450 Jahre Kroaten im Burgenland/450 ljet Hrvati u Gradišću: Katalog der Sonderausstellung des Burgenländischen Landesarchivs*. Amt der Burgenländischen Landesregierung: Eisenstadt. 1983a.

Tobler, Felix: "Die kroatische Kolonisation im Wieselburger Komitat im 16. Jahrhundert". *Burgenländische Heimatblätter* 45, 1983b, pp. 119–142.

Tobler, Felix: "Herkunft und Ansiedlung". In: Geosits, Stefan (ed.): *Die burgenländischen Kroaten im Wandel der Zeiten*. Edition Tusch: Wien. 1986, pp. 3–50.

Tobler Felix: "Kroaten und Deutsche in Kittsee. Beispiele zu ihrem Verhältnis im 18. und 19. Jahrhundert". *Burgenländische Heimatblätter* 54, 1992, pp. 18–25.

Tornow, Siegfried: " Vom Kroatischen zum Deutschen. Etappen des Sprachenwechsels im Burgenland". In: Holzer, Werner/Münz, Rainer (eds.): *Trendwende? Sprache und Ethnizität im Burgenland*. Passagen: Wien. 1993, pp. 149–154.

Vašková, Lenka: *Rozsídlení moravských Chorvatů a postoj jugoslávské vlády*. (BA thesis) Masarykova univerzita: Brno. 2013. retrieved 12.01.2019, from https://is.muni.cz/th/bh99i/Bakalarska_prace.pdf

Weilguni, Werner: *Die Diskussion um die Standardsprache bei den Burgenlandkroaten. Sprachpolitische Tendenzen seit der Mitte des 19. Jahrhunderts*. (doctoral thesis) Universität Wien: Vienna. 1984.

Anna Tetereva, Viktoria Naukhatskaia

Variation im Spracherwerb von Verben bei bilingualen Kindern (Russisch – Deutsch)

Abstract: The present contribution reports a study on the parallel development of verbs and verbal categories in Russian and German in sequentially bilingual children. The data of eight Russian-German bilingual children, age 6–7, were elicited in two consecutive years (2016 and 2017) by using the Multilingual Assessment Instrument for Narratives (MAIN). The study focused on the development of the verbal system in L1 (Russian) and L2 (German) in bilingual children by comparing the usage of verbs and verbal categories in language production. Furthermore, the study evaluates how L1 and L2 affect each other in language production of bilinguals.

Keywords: bilingualism, Russian, German, verbal categories, language production, language acquisition

1 Einleitung[1]

Gegenwärtig ist Russisch die größte Migrantensprache in Berlin und die Herkunftssprache vieler Kinder, die diese Sprache gleichzeitig mit der deutschen Sprache erwerben, d. h. bilingual aufwachsen. In Anlehnung an die Daten aus dem Bericht des Berliner Interdisziplinären Verbunds für Mehrsprachigkeit (BIVEM) aus den Jahren 2011–2014 gehören laut Gagarina (2013, S. 197) gerade Kinder mit Migrationshintergrund, die zwei- oder mehrsprachig aufwachsen, zu der Gruppe, die in Bezug auf das Deutsche oft einen Leistungsrückstand bzw. mangelnde Sprachkompetenzen aufweist. Es wurden im schulischen und vorschulischen Bereich zahlreiche Sprachfördermaßnahmen für mehrsprachige Kinder ausgearbeitet, die aber oft nicht den gewünschten Erfolg zeigen. Ein

---

Grund hierfür ist, dass in der erzieherischen Praxis zu wenig über die Grundlagen des bilingualen Spracherwerbs bei Kindern mit unterschiedlichen Erstsprachen bekannt ist (vgl. Gagarina et al. 2014, S. 139).Vor diesem Hintergrund verwundert es nicht, dass in Berlin Bilingualismus und Mehrsprachigkeit Forschungsschwerpunkte am Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS) sind, wo bereits umfangreiche Studien zum bilingualen Spracherwerb Russisch/Deutsch erarbeitet wurden (vgl. z. B. Gagarina 2009, 2011, 2013; Klassert 2011). Zudem erstellt das ZAS im Rahmen des BIVEM ein Korpus aus Sprachdaten in der Herkunfts- (Russisch) und Umgebungssprache (Deutsch), die mit Hilfe des Multilingual Assessment Instruments for Narratives (MAIN) (vgl. Gagarina et al. 2012, 2015) gesammelt wurden. Durch MAIN, einem Test zur Erzählfähigkeit von monolingualen und bilingualen Kindern, werden das Verstehen und die Produktion von Geschichten geprüft.

Der vorliegende Artikel ist eine Studie zur parallelen Entwicklung von Verben und Verbalkategorien im Russischen und im Deutschen bei sequenziell bilingualen Kindern. Analysiert werden die mit Hilfe von MAIN erhobenen Daten von acht Kindern im Alter von sechs und sieben Jahren aus zwei aufeinanderfolgenden Jahren (2016–2017), um folgende Forschungsfragen zu beantworten:

1. Wie entwickelt sich das Verbsystem bei bilingualen Kindern in der L1 (Erstsprache) und L2 (Zweitsprache)?
2. Zeigen die Kinder ähnliche Muster im Gebrauch von Verben und Verbalkategorien in der Sprachproduktion in beiden Sprachen?
3. Wie beeinflusst die L2 die Produktion von Verben in der L1 und umgekehrt?

Dabei bieten die Narrative von Kindern eine Reihe von Vorteilen als Einstiegspunkt für die vorliegende Studie. Einerseits ermöglichen sie einen Blick auf mehrere sprachliche Phänomenbereiche, wobei die folgenden in Bezug auf das Verb fokussiert wurden: Lexik (lexikalische Vielfalt von Verben, präfigierte Verben), Morphosyntax (Tempus, Aspekt) und mit Bilingualismus in Zusammenhang stehende Phänomene sprachübergreifender Übertragung). Andererseits ist die Struktur von Kindererzählungen relativ unveränderlich und unabhängig von der Produktionssprache (vgl. Iluz-Cohen/Walters 2012), was den zwischensprachlichen Vergleich erleichtert.

2 Sequenzieller Bilingualismus, Spracheinfluss und Sprachdominanz

Da der Gegenstand dieser Studie bilinguale Kinder sind, sollen zunächst der Begriff Zweisprachigkeit als Spezialfall der Mehrsprachigkeit definiert

und unterschiedliche Formen von Zweisprachigkeit aufgezeigt werden. Mehrsprachigkeit meint, dass sich ein Individuum in zwei oder mehr Sprachen verständigen kann. Bilingualismus – Zweisprachigkeit – ist die minimale Variante der Mehrsprachigkeit.

In Bezug auf Bilingualismus im frühen (kindlichen) Spracherwerb wird zwischen simultanem und sequentiellem Zweitsprachenerwerb unterschieden. In beiden Fällen handelt es sich um ein natürliches (ungesteuertes) Erlernen der Sprachen. Von simultanem Zweitsprachenerwerb spricht man, wenn Kinder innerhalb der ersten drei Lebensjahre gleichzeitig mit zwei Sprachen konfrontiert werden. Sequentiell zweisprachig sind hingegen diejenigen Kinder, die zunächst eine und danach eine zweite Sprache entwickeln (vgl. Belliveau 2002, S. 15). Sprechen beide Elternteile oder zumindest eine andere regelmäßige Bezugsperson, wie zum Beispiel ein Kindermädchen, eine andere Sprache, so beginnt der simultane Spracherwerb mit der Geburt. Auf diese Weise ist das Kind in der Lage, zwei Sprachen gleichzeitig zu erwerben und bis zur Einschulung auf dem Niveau eines einsprachigen Muttersprachlers zu beherrschen. Für die Entwicklung des Bilingualismus spielen auch ethnolinguistische Aspekte eine große Rolle. Dabei werden bi-ethnischer und mono-ethnischer Bilingualismus (vgl. Tchirsheva 2012, S. 65) unterschieden.

Nach Tchirsheva (2012) kann man von simultanem Bilingualismus nur dann sprechen, wenn das Kind mit der L2 vor der aktiven Verwendung von Wörtern der L1 (d. h. nicht später als im Alter von 10 Monaten) konfrontiert wird. Je später die zweite Sprache in der Kommunikation mit dem Kind verwendet wird, desto offensichtlicher erkennt man Merkmale der Sprachdominanz in Morphologie, Lexikon, Syntax, Semantik und Pragmatik. Die zweite Sprache ist hinsichtlich des Grades ihrer Beherrschung der Erstsprache somit untergeordnet. Fast immer ist eine der beiden Sprachen schwächer ausgebildet als die andere, was von verschieden Faktoren abhängt. Wesentlich sind natürlich vor allem die Sprachverteilung innerhalb von Familie und sozialem Umfeld oder das Land, in welchem sich diese Menschen gewöhnlich aufhalten. Ändern sich einige oder gar alle genannten Faktoren, kann das auch Auswirkungen auf die Beherrschung der beiden Sprachen haben, sich die Differenz vergrößern, verringern oder das Verhältnis der beiden Sprachen zueinander auch umkehren (vgl. Leist-Villis o.J.). Sprachdominanz ist somit von Hintergrundfaktoren (*background factors*, vgl. Gagarina 2017) beeinflusst.

Attrition als eine besondere Form des Dominanzwechsels bedeutet den schrittweisen Verlust einer Sprache bei einem Individuum (Schmid 2002, S. 7). Dabei ersetzt die neu erlernte L2 die L1 und verdrängt deren Dominanz. Die Angehörigen der zweiten Migrationsgeneration, also Personen, die in

Deutschland geboren wurden oder als Kinder mit ihren Eltern kamen, erwerben die L1 nicht nur in unterschiedlichem Maße, sondern bewahren dieses Wissen auch in verschiedenem Umfang – häufig kommt es nach der frühen Kindheit zur Attrition bereits erworbener Strukturen.

Man geht davon aus, dass die beiden Sprachen in einem gemeinsamen Repräsentationssystem existieren und miteinander agieren (vgl. Klassert 2011). Borrowings werden im Zusammenhang mit Bilingualität als Ausweichstrategie (Relief-Strategie) beschrieben (vgl. Deuchar/Quay 1999): Ein Gesprächspartner bedient sich Lexemen der L1, da für das geforderte Wort in der L2 keine geeignete Lexikalisierung zur Verfügung steht. Im einsprachigen Modus, in einer Situation also, in der die bilinguale Person mit einer monolingualen Person kommuniziert, treten Borrowings im Allgemeinen nicht auf, vielmehr wird das Einfügen eines Lexems in einer solchen Situation auf eine Interferenz, das Durchsetzen eines Wortes der falschen Sprache im Aktivierungsprozess, zurückgeführt (vgl. Van der Linden 2000). Für russisch-deutsche Kinder wurde durch Anstatt/ Dieser (2007) gezeigt, dass Mischäußerungen in der monolingualen russischen Kommunikation älterer bilingualer Kinder (4–5, 8–9 Jahre) häufiger sind als in der deutsch-monolingualen Kommunikation. Naheliegend schließen die Autoren, dass dieses Phänomen darauf zurückzuführen ist, dass die Kinder in monolingual-deutscher Kommunikation geübter sind.

In Abgrenzung zum kollektiven Phänomen des Borrowing, wird die Interferenz als individuelle Erscheinung definiert, die durch die Übertragung von Strukturen einer Sprache in die andere Sprache gekennzeichnet ist. Inhaltlich eng verwandt ist der in der Mehrsprachigkeitsforschung verwendete Begriff der Transferenz, der die Übertragung von Regeln und Merkmalen einer Sprache in eine andere Sprache bezeichnet. Solche Prozesse können das Erlernen einer Fremdsprache positiv, aber auch negativ beeinflussen.

Anstatt/Dieser (2007, S. 160) verweisen außerdem darauf, dass Sprachmischungen in beiden Sprachen in ihren Daten in den meisten Fällen auf „Wortnot" zurückzuführen waren, also als Borrowing und nicht als Interferenz zu interpretieren sind.

2 Aspekte einer mikrostrukturellen Analyse

Im Rahmen dieser Untersuchung konzentrieren wir uns auf die so genannte mikrostrukturelle Analyseebene (vgl. Gagarina et al. 2012). Diese umfasst eine breite Palette von linguistischen Aspekten, wie die Länge der Aussagen, lexikalische Vielfalt, morphosyntaktische Aspekte und Zweisprachigkeitsphänomene. Dabei sind die Mikrostrukturelemente sprachspezifisch und es ist entsprechend

unvermeidlich, dass sie auch im Russischen und Deutschen unterschiedlich sind (vgl. ebd.: 55). Im Folgenden wird ergänzend zu den weiter oben schon erwähnten Aspekten des Bilingualismus (sprachübergreifende Übertragung) in Bezug auf das Verb kurz auf die für diese Studie relevanten Ebenen der Lexik (lexikalische Vielfalt von Verben) und der Morphosyntax (Tempus, Aspekt) eingegangen, um einen Interpretationshintergrund für die im Anschluss beschriebenen Ergebnisse zu bilden.

Die lexikalische Vielfalt und insbesondere der anteilsmäßige Erwerb von Nomen und Verben durch bilinguale Kinder wird von Klassert (2011, S. 98) als möglicher Indikator für die Dominanz einer bestimmten Sprache identifiziert. Die Ergebnisse der wenigen vorliegenden Studien, die von Klassert (2011, S. 99–101) beschrieben wurden, sind dabei nicht eindeutig. Einige Studien (Karasu 1995; Hepsöyler/Liebe-Harkort 1991; David/Wei 2005; Sheng et al. 2006) zeigen, dass für beide Wortarten eine parallele Entwicklung in beiden Sprachen stattfindet und dass sich der Status einer Sprache nicht unterschiedlich auf den Erwerb von Nomen und Verben auswirkt. Im Gegensatz dazu legen die Ergebnisse von anderen Studien nah (Jia et al. 2006; Kohnert et al. 1999), dass die Entwicklung beider Wortarten vom Status einer Sprache beeinflusst sein kann. So kommt es mit steigendem Alter zu einer Verschiebung der Sprachdominanz. Die stärkeren Fähigkeiten in L1 bei den Fünf- bis Siebenjährigen wurden durch stärkere Fähigkeiten in L2 bei den älteren Gruppen abgelöst. Es wurde auch festgestellt, dass die lexikalischen Fähigkeiten für Verben in L2 die Fähigkeiten in L1 früher als für Nomen übersteigen (Klassert 2011, S. 160). Die Studie von Klassert (2011) hat auch gezeigt, dass bei den Nomina bereits im Grundschulalter ein Sprachabbau in L1 beobachtbar ist, wenn die Kinder in dieser Sprache nicht gefördert werden. Bei den Verben kann jedoch ein geringes Fortschreiten in der Entwicklung beobachtet werden. Im Gegensatz dazu steigen die Benennfähigkeiten der bilingualen Kinder im Deutschen in der dominanten Sprache mit dem Alter für Nomen und für Verben an (vgl. Klassert 2011, S. 160). Auf Grund der oben dargestellten Befunde lässt sich keine eindeutige Hypothese dahingehend aufstellen, ob eine ähnliche Situation in den erzählten Geschichten von russisch-deutschen bilingualen Kindern in Bezug auf die lexikalische Vielfalt der Verben zu beobachten sein wird.

Auf der Ebene der Morphosyntax sind in der vorliegenden Studie die Tempus- und die Aspektkategorie von besonderem Interesse. Der Verbalaspekt gehört zu denjenigen sprachlichen Erscheinungen, in denen sich das Russische besonders auffällig vom Deutschen unterscheidet, da das Deutsche über diese grammatikalische Form nicht verfügt. Auch im Erwerb des Russischen als Herkunftssprache ist der Aspekt eine besondere Schwierigkeit (Anstatt 2008). Im Russischen werden zwei Aspekte unterschieden:

1. Perfektiv (vollendeter Aspekt)
2. Imperfektiv (unvollendeter Aspekt)

Anstatt (2008) untersucht die Verwendung von Aspekt und Tempus in russischen und deutschen Erzählungen in Form von Bildergeschichten bei mono- und bilingualen Kindern im Alter von drei bis neun Jahren. Sie zeigt, dass monolinguale Kinder sprachabhängige Erzählstrategien bevorzugen. Ein erheblicher Anteil der bilingualen Kinder nutzt das Tempus (im Deutschen) bzw. Aspekt und Tempus (im Russischen) anders als monolinguale Kinder, vermutlich weil diese Kinder auf Erzählstrategien der anderen Sprache zurückgreifen. Aus den Ergebnissen ihrer Untersuchung ist ersichtlich, dass die bilingualen Probanden in den Erzählungen im Russischen das Präsens, das imperfektive und perfektive Präteritum benutzt haben, z. B. *padaet* – *padal* – *upal* ´fällt – fiel – gefallen´. Im Deutschen benutzten sie hingegen das Präsens, Präteritum und Perfekt, z. B. *fällt herunter* – *fiel herunter* – *ist heruntergefallen*. Von monolingualen russischen Kindern werden Geschichten bevorzugt im Präteritum erzählt, von monolingualen deutschen Kindern hingegen im (historischen) Präsens. Was bilinguale Kinder anbelangt, so zeigt sich, dass die Hälfte der bilingualen Kinder in ihren russischen Erzählungen das Präsens bevorzugt. Außerdem übertragen einige wenige das russische Modell – die Erzählung im Präteritum – ins Deutsche und verwenden hier das Perfekt. Nach Anstatt (2017) stellt der Erwerb und Erhalt des Verbalaspekts eine besondere Herausforderung dar. Außerdem kann es bei den Personen, die das Russische nur in einem sehr geringen Umfang erwerben bzw. bewahren zum vollständigen Schwund oder Nichterwerb der Kategorie kommen. An dieser Stelle ist es wichtig zu betonen, dass nach Gagarina (2017, S. 401) auch der Abbau bzw. die Erosion von bereits erworbenen Elementen stattfinden kann, wie sie sich z. B. in Kasusfehlern zeigen.

3 Daten

Die vorliegende Untersuchung basiert auf Material, das mit Hilfe des Tests zur Erzählfähigkeit von monolingualen und bilingualen Kindern (MAIN: Multilingual Assessment Instrument for Narratives, vgl. Gagarina et. al. 2012) erhoben wurde.

   Durch MAIN werden das Verstehen und die Produktion von Geschichten bei Kindern der Altersgruppe zwischen drei und zehn Jahren geprüft. Das Testmaterial besteht aus vier farbigen Bildergeschichten mit jeweils sechs Bildern, die erzählt bzw. nacherzählt werden sollen. Die Geschichten sind so konzipiert, dass sie die Produktion von bestimmten Mikrostrukturen fordern,

z. B. die Nutzung von bestimmten Verben oder Konnektoren. Das ermöglicht die Messung des Sprachstandes der Kinder in beiden Sprachen. Dabei wird das Erzählen (mündliches bzw. schriftliches) als Testformat ausgewählt, da dieses im Vergleich zum Nacherzählen den Kindern mehr Freiheit bietet, ihre eigene Vorstellungskraft zu benutzen. Dies ist seinerseits für die Bewertung von Kindersprache von Bedeutung, weil darüber nähere Informationen über unabhängige narrative Formulierungsfähigkeiten der Kinder gewonnen werden können (vgl. Gagarina et. al. 2012).

Diese Studie bezieht sich auf 38 Texte von acht Probanden, die mit Hilfe von MAIN in den Jahren 2017 und 2016 erhoben wurden: 16 Transkripte von diktierten Geschichten aus dem Jahr 2016 und 16 handgeschriebene Texte aus 2017. Die letzteren wurden von uns entziffert, transkribiert, analysiert und mit den Daten aus dem Jahr 2016 verglichen.

Zum Gegenstand unserer Untersuchung liegen bereits umfangreiche Daten russisch-deutscher Kinder in verschiedenen Altersgruppen vor (BIVEM, 2011–2014). Für die hier vorgenommenen Analysen wurden nur die Daten von Kindern im Alter von sechs bis acht Jahren verwendet. Als Datenbasis für die vorliegende Untersuchung dienen die Sprachdaten von acht Kindern. Die Daten der bilingual russisch-deutschen Kinder wurden in Berliner Schulen erhoben. Entwicklungsstörungen wurden anamnestisch ausgeschlossen.

Das zweisprachige Korpus wurde nach den folgenden probandenbezogenen, soziolinguistischen Kriterien zusammengestellt, um die Input- und Erwerbssituation zwischen den Kindern möglichst homogen zu halten. Die folgenden Daten wurden im Rahmen einer Befragung der Eltern der Kinder erhoben, die im Anschluss hinsichtlich ihrer zweisprachigen Entwicklung evaluiert wurden:

1. Alle Kinder wurden in Deutschland geboren, leben seit ihrer Geburt in Deutschland und wachsen mit Russisch als Familiensprache auf. Sechs der zehn Kinder in unserem Sample verwenden in der Familie neben dem Russischen auch das Deutsche. Den Elternfragebögen zufolge wurden die beiden Sprachen in verschiedener Reihenfolge erworben: Sechs Kinder haben zuerst Russisch als Familiensprache erworben und sind somit sequentiell-bilingual, vier Kinder sind nach Angaben von Eltern von Geburt an zweisprachig.
2. Alle Kinder besuchten spätestens seit ihrem dritten Geburtstag den Kindergarten. Die meisten, nämlich sieben der Kinder, verbrachten sechs bis sieben Stunden pro Tag in der Kita. In einigen Kitas gab es laut Aussage der Eltern viele Kinder, die Russisch sprachen.

3. Beide Elternteile der Kinder sind Migranten der ersten Generation. Die Eltern von einem der zehn Kinder leben seit über zwanzig Jahren in Deutschland, die von vier seit über zehn Jahren und die der fünf weiteren seit über fünf Jahren.

4. Fast alle Mütter (sieben von zehn) haben einen mittleren Bildungsabschluss, sechs Mütter üben derzeit keinen Beruf aus und sind Hausfrauen, zwei machen eine Ausbildung, eine ist als einfache Fachkraft tätig. Eine Mutter machte zum Zeitpunkt der Datenerhebung eine schulische Ausbildung. Über die Berufstätigkeit der Väter gibt es in den meisten Fällen keine Angabe.

5. Die Eltern sprechen zu Hause mit dem Kind überwiegend Russisch. Vier der Familien geben an, dass sie zu Hause eine Mischsprache (Russisch und Deutsch) benutzen, zwei wechseln zwischen den beiden Sprachen, in einer Familie wurden Russisch und eine Mischsprache benutzt und nur drei Familien sprechen zuhause ausschließlich Russisch. Sieben Kinder sprechen nach Einschätzung der Eltern zuhause und meist nur mit den Eltern Russisch. Mit ihren Freunden sprechen fünf Kinder Russisch, drei Kinder sprechen überall Russisch. Außerhalb der Kita sprechen drei Kinder auch zuhause Deutsch (meistens mit den Geschwistern), fünf mit ihren Freunden und ein Kind spricht überall Deutsch. Nur zwei Mütter sprechen mit ihrem Kind ausschließlich Russisch, acht sprechen Russisch und Deutsch (jedoch viel Russisch und wenig Deutsch). Die Mütter, die mit ihren Kindern ausschließlich Russisch sprechen, haben geringe Deutschkenntnisse. Fünf Mütter geben an, dass sie gut Deutsch sprechen.

6. Fast allen Eltern ist es wichtig, dass ihr Kind Russisch nicht vergisst. Neun Eltern lesen den Kindern auf Russisch vor, acht versuchen, so viel Russisch wie möglich zu sprechen, sieben benutzen russische TV-Sendungen als Mittel für die Sprachförderung, zwei machen Übungen und sechs Eltern halten die Kinder aktiv dazu an, Russisch zu sprechen. Den Elternfragebögen zufolge sprechen sechs Kinder lieber Russisch als Deutsch, zwei bevorzugen Deutsch und zwei Kinder sprechen beide Sprachen gleich gern. Außerdem wurde bei der Auswahl der Probanden darauf geachtet, dass die Menge des deutschen bzw. russischen Inputs in der Bildungseinrichtung vergleichbar ist. So wurden keine Kinder in die Studie aufgenommen, die Kindergärten mit bilingualem Bildungsprogramm besuchten.

5 Ergebnisse

5.1 Datenanalyse: Deutsch

Insgesamt wurden 16 deutschsprachige Texte analysiert (s. Tab. 1). Das deutschsprachige Datenkorpus aus dem Jahr 2017 umfasst insgesamt 401

Tab. 1: Datenvergleich Verb 2016–2017 (Deutsch)

Deutsch	MAIN 2017 (schriftlich)	MAIN 2016 (mündlich)
Tokens	401/50 (24/68)	794/100 (22/178)
Verben pro Text	68/9 (6/13)	155/18 (8/36)
Hapax Legomina	62/7 (5/11)	91/12 (5/20)
Präfigierte Verben	13/2 (0/3) *weg-* (4), *auf-* (3)	36/5/ (2/10) *weg-* (11), *auf-* (7), *raus-* (5), *zu-* (3)
Zeitformen	Präsens: 32 Verben Präteritum: 7 Verben Perfekt: 22 Verben	Präsens: 58 Präteritum: 17 Perfekt: 41
Hauptform der Erzählung	Perfekt: 4 Probanden Präsens: 3 Probanden Präteritum: 1 Proband	Perfekt: 4 Probanden Präsens: 3 Probanden keine Hauptform: 1 Proband

Tokens.[2] Wenn man die Verteilung von Verben in diesem Korpus betrachtet, so sind 68 aus 401 Tokens Verben, die anderen 333 sind andere Wortarten inklusive Determinatoren (bestimmte und unbestimmte Artikel), die in den russischen Texten nicht präsent sind.

Die Durchschnittslänge der zehn analysierten, auf Deutsch verfassten Geschichten aus der Erhebungsrunde 2017 beträgt 50 Tokens. Der längste Text umfasst 68 Tokens, der kürzeste 24. Dabei enthalten die Texte durchschnittlich neun Verben (minimal sechs und maximal 13). Darunter sind durchschnittlich sieben Hapax legomina (minimal sechs und maximal elf). Durchschnittlich sind pro Text zwei Verben präfigiert. Am häufigsten werden Verben mit den Präfixen *weg-* (6) und *auf-* (3) benutzt. Andere Präfixe wurden jeweils nur einmalig im gesamten Korpus gebraucht.

Betrachtet man die Texte aus der Erhebungsrunde 2016, in der die Geschichten nicht schriftlich beschrieben, sondern mündlich erzählt wurden, kommt man zu folgenden Ergebnissen: Im Vergleich zu den schriftlichen Texten umfasst das Korpus der deutschsprachigen Texte mehr, nämlich 794 Tokens, darunter sind 155 Tokens Verben und 639 weitere Wortarten (vgl. Abb. 1). Die Durchschnittslänge der mündlichen Texte liegt bei 100 Tokens pro Text. Maximal wurden die Texte mit 178 und minimal mit 22 Tokens realisiert. Dabei enthalten die Texte durchschnittlich 18 Verben (minimal acht und maximal

---

2   Als Token wurden alle im Text verständlich produzierten Wortarten betrachtet, die zusammengesetzten Perfektformen von Verben wurden als ein Token gezählt.
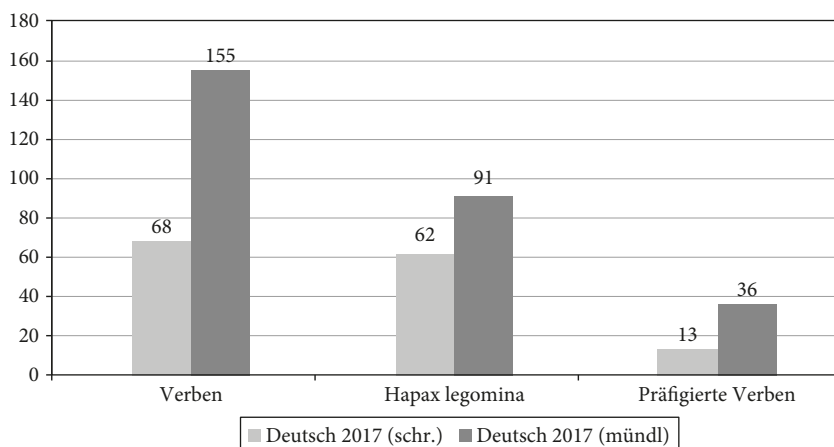
Abb. 1: Verbgebrauch im deutschsprachigen Datenkorpus 2016–2017

36). Darunter sind durchschnittlich zwölf Hapax legomina (minimal fünf und maximal 20). Durchschnittlich werden pro Text drei präfigierte Verben benutzt, wobei am häufigsten Verben mit den Präfixen *weg-* (11), *auf-* (7), *raus-* (5), *zu-* (3) auftreten. Andere Präfixe (*an-, rein-, an-, hoch-, mit-, runter-*) finden sich im gesamten Korpus nur einmal. Im Vergleich zu den Daten aus dem Jahr 2017 zeigt sich, dass das Korpus aus dem Jahr 2016 in Bezug auf die Anzahl der Token und Verben sowie die Anzahl der präfigierten Verben beinahe doppelt so groß ist, wie jenes aus dem Jahr 2017. Dies lässt sich mit der Wahl des Erhebungsmodus (schriftlich vs. mündlich) erklären. Auf das Verhältnis von Verben und anderen Wortarten im Gesamtkorpus hat der Erhebungsmodus, wie Abb. 2 zeigt, offenbar keinen Einfluss gehabt.

Bei der Analyse der temporalen Struktur wurden alle finiten Verben in den jeweiligen Zeitformen aus den Texten gezählt.[3] Dabei sind wir zu folgenden Ergebnissen gelangt: In den deutschsprachigen Texten wurden 52 % der Verben im Präsens, 36 % der Verben im Perfekt und 11 % der Verben im Präteritum verwendet (vgl. Abb.3).

Es ist zu beachten, dass sich der Gebrauch der Verben in einer bestimmten Zeitform nicht auf einen produzierten Text bezieht. Um die Beschreibung

---

3   In Konstruktionen mit Modalverben, z. B. *wollte trinken*, wurde nur das Modalverb *wollte* berücksichtigt.
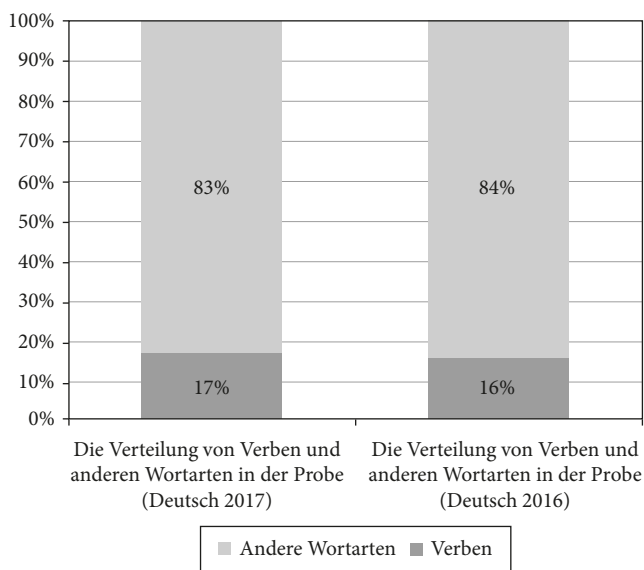
Abb. 2: Verteilung von Verben und anderen Wortarten im deutschsprachigen Korpus 2016–2017

der temporalen Struktur zu präzisieren, wurde in allen Texten die temporale Hauptform pro Proband bestimmt (vgl. Abb. 4). Wie Anstatt (2008: 11) verstehen wir unter der temporalen Hauptform die Zeitform, die bei 50 % aller finiten Verben gebraucht wird. So haben im Jahr 2017 vier Probanden das Perfekt, drei Personen Präsens und eine Person das Präteritum als Hauptform des Erzählens gewählt.

Genauso wie in den Texten aus dem Jahr 2017 ist die Zeitformenverteilung pro Text quantitativ fast gleichgeblieben. So wird von vier Probanden das Perfekt, drei Probanden das Präsens und von einem Probanden das Präteritum als Hauptform bevorzugt. Diese Ergebnisse stimmen allerdings mit den Ergebnissen von bilingualen Kindern aus der Studie von Anstatt (2008, S. 16) nicht überein, denen zufolge 74 % der bilingualen Kinder das Präsens und nur 24 % das Perfekt als Hauptform der Erzählung wählen.
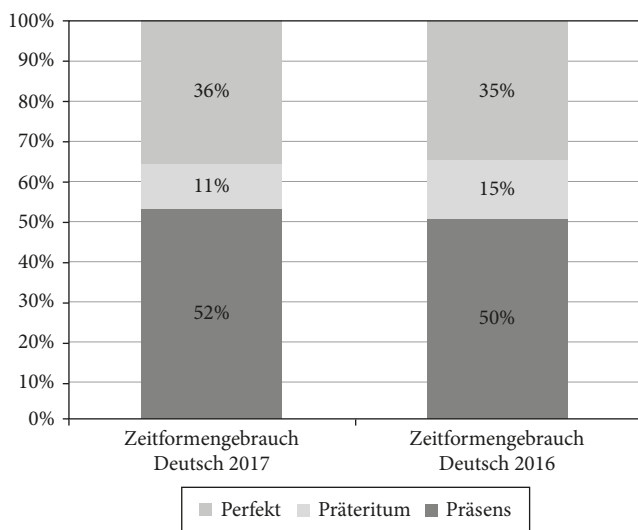
Abb. 3: Die Verteilung der Zeitformen im deutschsprachigen Datenkorpus 2016–2017

Bei der Textanalyse im lexikalischen Bereich war die Beschreibung einer bestimmten Handlung auf dem letzten Bild in beiden Geschichten von besonderem Interesse. Es handelt sich um das Verb *essen/fressen* und dessen präfigierte Formen in Bezug auf die Handlung des Hundes bzw. der Katze, die auf dem Bild Würstchen bzw. Fische auffressen. In zehn Texten wurden fünf verschiedene Verben gebraucht:

- *essen* (3)
- *aufessen* (3)
- *fressen* (1)
- *sich vollfressen* (1)
- *wegfressen* (1)

Der Gebrauch der Verben *essen* und *aufessen* könnte als sprachkontaktinduziert erklärt werden. Während sich im Russischen das Äquivalent für das Verb *essen* sowohl auf eine Person als auch auf ein Tier bezieht, wird im Deutschen bei Tieren primär *fressen* gebraucht. In den deutschen Texten der Probanden sehen wir, dass *fressen* und seine präfigierten Formen nur dreimal gebraucht wurden. Es könnte sein, dass es sich in diesem Fall um eine Interferenz handelt. Dabei ist auffallend, dass keine weiteren Interferenzphänomene im lexikalischen Bereich festgestellt wurden.
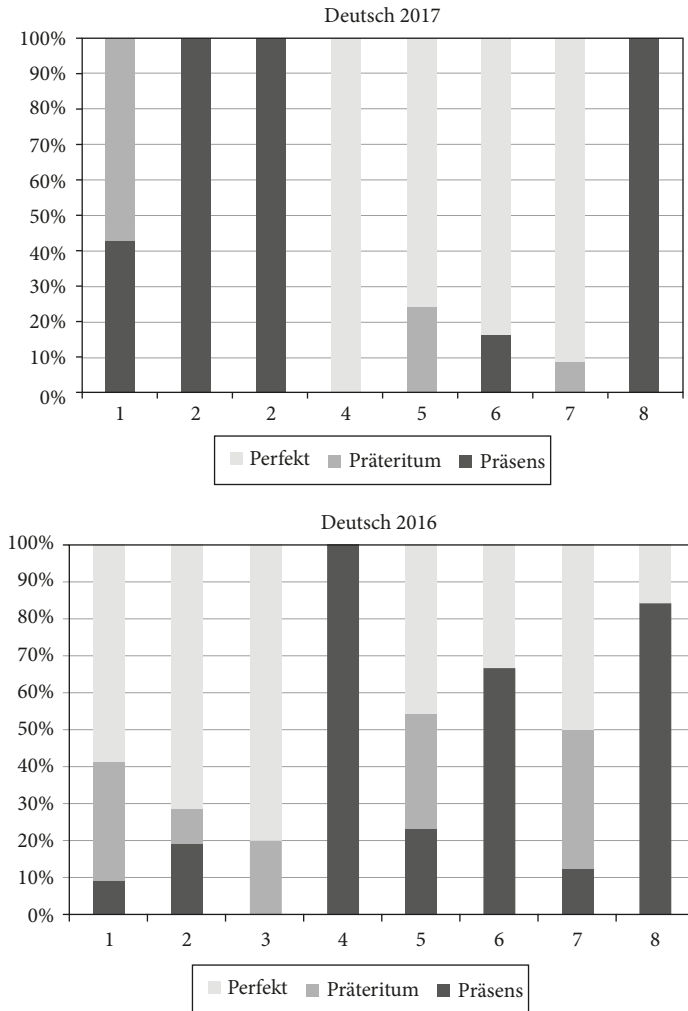
Abb. 4: Verteilung des Zeitformengebrauchs pro Proband im deutschsprachigen Datenkorpus 2016–2017

Tab. 2: Datenvergleich Verb 2016–2017 (Russisch)

Russisch	MAIN 2017 (schriftlich) insgesamt/durchschnittlich (minimal/maximal) pro Text	MAIN 2016 (mündlich) insgesamt/durchschnittlich (minimal/maximal) pro Text
Tokens	248/26 (21/50)	448/56 (30/106)
Verben	66/8 (5/21)	111/14 (7/27)
Hapax Legomina	58/7 (min.4, max.11)	84/11 (min.6, max. 17)
Präfigierte Verben	27/4 (min. 1, max. 6): *u-* (7), *pri-* (4), *s-* (3), *vy-* (3), *po-* (2), *na-* (1), *za-* (1)	47/6 (min. 3, max. 10): *u-* (27), *vy-* (6), *po-* (8), *pri-* (4), *s(o)-* (5), *za-* (5)
Zeitformen	Präsens 14 Verben Präteritum (Perf.) 46 Verben Präteritum (Imp.) 8 Verben	Präsens 19 Verben Präteritum (Perf.) 54 Verben Präteritum (Imp.) 27 Verben
Hauptform der Erzählung	Präsens 0 Probanden Präteritum (Perf.) 8 Probanden Präteritum (Imp.) 0 Probanden	Präsens 1 Proband Präteritum (Perf.) 5 Probanden Präteritum (Imp.) 1 Proband keine Hauptform 1 Proband

5.2 Datenanalyse: Russisch

Insgesamt wurden 16 russischsprachige Texte analysiert (s. Tab. 2). Das russischsprachige Datenkorpus aus dem Jahr 2017 umfasst 248 Tokens. Wenn man die Verteilung von Verben in diesem Korpus betrachtet, so sind 66 Tokens Verben (vgl. Abb. 6).

Die Anzahl der Tokens in den auf Russisch verfassten Texten aus dem Jahr 2017 liegt bei durchschnittlich 26 (minimal 21, maximal 50). Die Gesamtanzahl von Verben pro Text liegt bei durchschnittlich acht (minimal fünf, maximal zwölf). Darunter sind durchschnittlich sieben Hapax legomina (minimal vier, maximal elf) (vgl. Abb. 7).

Außerdem wurden die Anzahl und morphologische Struktur der präfigierten Verben analysiert. In den Texten gibt es durchschnittlich vier präfigierte Verben (minimal eines, maximal sechs). Meistens verwenden die Probanden folgende Präfixe: *u-* (7), *pri-* (4), *s-* (3), *vy-* (3), *po-* (2), *na-* (1), *za-* (1)

Auf die gleichen Kriterien haben wir bei der Analyse der Texte aus dem Jahr 2016 geachtet. Die Anzahl der Tokens liegt bei durchschnittlich 56 (min. 21, max. 50). Die Gesamtanzahl von Verben pro Text liegt bei durchschnittlich 14 (min. sieben, max. 27). Darunter finden sich durchschnittlich elf Hapax legomina (min. sechs, max. 17).

Die temporale Struktur der schriftlichen Texte ist durch den Gebrauch von folgenden Zeitformen gekennzeichnet: Im gesamten Korpus der
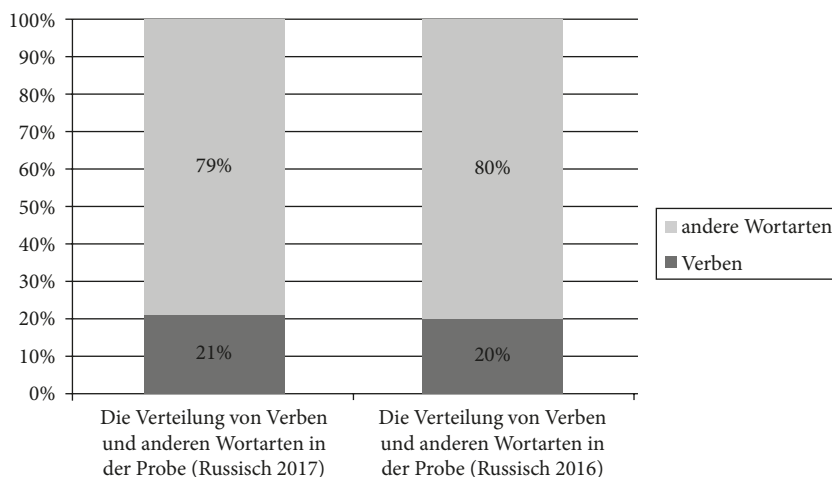
Abb. 5: Verteilung von Verben und anderen Wortarten im russischsprachigen Datenkorpus 2016–2017

russischsprachigen Texte aus dem Jahr 2017 finden sich 14 Verben im Präsens, 46 Verben im Präteritum perfekt und 8 Verben im Präteritum imperfekt. Sieben Probanden wählten das Präteritum und nur ein Proband das Präsens als Haupterzählform. Dabei haben sieben Probanden beim Gebrauch des Präteritums die perfektive Form gegenüber der imperfektiven bevorzugt.

Genauso wie in den Texten aus dem Jahr 2017 wurde als Hauptform des Erzählens das Präteritum von sieben Probanden ausgewählt, nur ein Proband benutzte das Präsens. Was die Verteilung der präteritalen Form angeht, so haben fünf Probanden primär die perfektive Form gebraucht und ein Proband die imperfektive Form (vgl. Abb. 9).

Wenn wir uns den Zeitformen zuwenden, dann sehen wir, dass unseren Ergebnissen zufolge aus dem gesamten Korpus der russischsprachigen Texte 19 Verben im Präsens, 54 Verben im Präteritum (pf.) und 27 Verben im Präteritum (ipf.) verwendet wurden (vgl. Abb. 8).

Präfigierte Verben wurden ebenfalls analysiert. In den Texten gibt es durchschnittlich sechs präfigierte Verben (minimal drei, maximal zehn). Am häufigsten verwenden die Probanden folgende Präfixe: *u-* (27), *vy-* (6), *po-* (8), *pri-* (4), *s(o)-* (5), *za-* (5).

Sowohl die Anzahl der Tokens und Verben als auch die der präfigierten Verben hat sich im Vergleich zu der Evaluation aus dem Jahr 2017 verdoppelt,
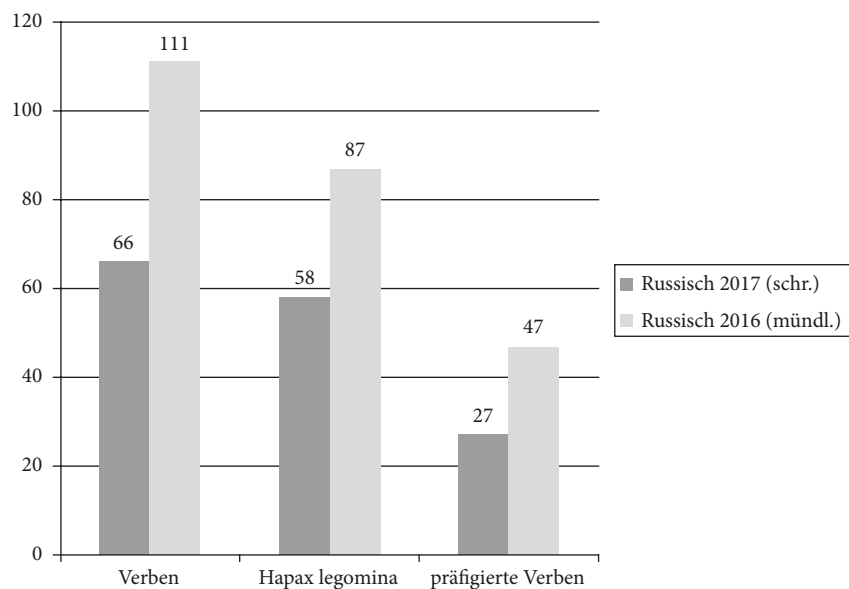
Abb. 6: Verbgebrauch im russischsprachigen Datenkorpus 2016–2017
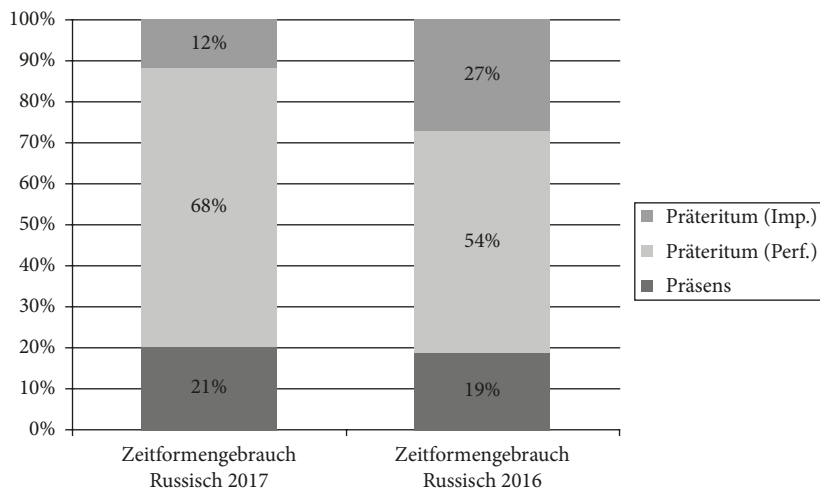
Abb. 7: Die Verteilung der Zeitformen im russischsprachigen Datenkorpus 2016–2017
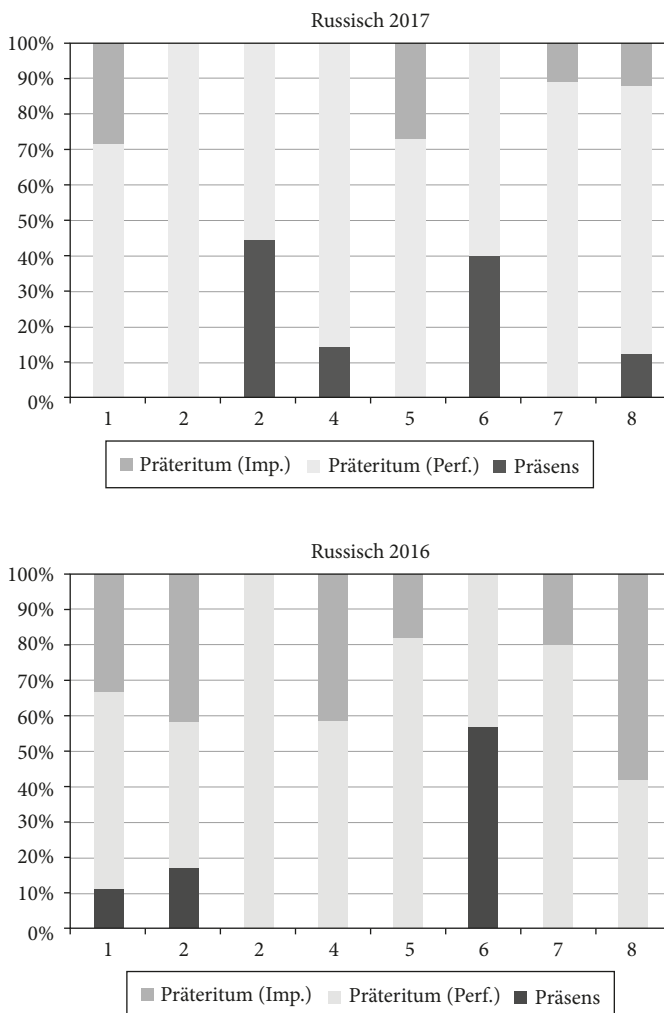
Abb. 8: Verteilung des Zeitformengebrauchs pro Person im russischsprachigen Datenkorpus 2016–2017

was jedoch vermutlich ganz wesentlich auf die Form der Evaluation (mündlich vs. schriftlich) zurückzuführen ist. Das Verhältnis von Verben und anderen Wortarten im russischsprachigen Korpus ist (auch wie in deutschsprachigen) aber auf vergleichbarem Niveau geblieben (vgl. Abb. 7).

Zudem gab es in den Texten aus dem Jahr 2017 einige Auffälligkeiten, die auf Interferenzen aus dem Deutschen hindeuten. So wurden bei der Analyse der Verbalphrasen einige Kasusfehler festgestellt, obwohl der Kasus im Alter von etwa sieben Jahren schon erworben sein sollte:

(1) a. *uvide-l   babašk-a.
            sehen.PFV-PST-M.SG   Schmetterling-NOM.SG
            '(Er) hat *der Schmetterling gesehen.'
      b. *uvide-l   kotk-a.
            sehen.PFV-PST-M.SG   Katze-NOM.SG
            '(Er) hat *die Katze (NOM) gesehen.'
      c. *košk-a   prygnu-ť.
            Katze-NOM.SG   springen.PFV-INF
            'Die Katze *springen.'

Auch haben wir in diesen Daten einige Beispiele gefunden, die sogenannte *Innovationen der Kindersprache* darstellen könnten. Zum Beispiel wurde das präfigierte Verb **spajmat'* 'fangen' (statt dem bereits perfektiven *pojmat'*) verwendet:

(2) *Mal'čik   *spajma-l*   mjač.
     Junge.NOM.SG fangen.PFV-PST-M.SG Ball.ACC.SG
     'Der Junge hat den Ball gefangen.'

Ein Kind wandte die inkorrekte präteritale Verbform **vazmi-l* 'nahm' (statt *vzja-l*) an:

(3) *Mal'čik   *vazmi-l*   mjačik.
     Junge.NOM.SG nehmen.PFV-PST-M.SG Ball.ACC.SG
     'Der Junge nahm das Bällchen.'

Auch die Daten aus dem Jahr 2016 zeigen bereits, dass die Kinder in diktierten Texten Verben sowohl im unvollendeten Aspekt als auch im vollendeten Aspekt anwenden, wie z. B. im folgenden Textausschnitt:

(4) I   potom   sobaka   potjanu-l-a   košku
    Und danach Hund.NOM.SG ziehen.PFV-PST.F.SG Katze.ACC.SG
    za xvost.
    am Schwanz. ACC.SG
    'Und danach hat der Hund die Katze am Schwanz gezogen.'
    I   potom   sobaka   *puga-l-a   košku.
    Und danach Hund.NOM.SG erschrecken.IPFV-PST-F.SG Katze.ACC.SG
    'Und danach erschreckte der Hund die Katze.'

Außerdem finden sich in unseren Daten einige Kasusfehler, entweder morphologisch als Indikatoren für den Erwerb des Kasussystems oder syntaktisch als solche, die für den Erwerb der lexemspezifischen Argumentstruktur betreffend interpretiert werden können.

(5) a. On   uvide-l   *odin   *ovečka.
         Er.NOM.SG   sehen.PFV-PST-M-SG   ein.NOM.M.SG   Schäfchen. NOM.F.SG
         'Er hat ein Schäfchen gesehen.'
     b. *Skazka   konec.
         Märchen.NOM.SG   Ende.NOM.SG
         'Das Märchen ist zu Ende.'⁴
     c. On   uvide-l   lisu,
         Er.NOM.SG   sehen.PFV-PST-M-SG   Fuchs.F.ACC.SG
         xotel   *ego   ukusi-t'.
         wollen.IPFV-PST-M-SG er.NOM.M.SG   beißen.INF
         'Er hat den Fuchs gesehen, er wollte ihn beißen.'
     d. Potom   koška   ubegaet   ot *sobaka.
         Dann   Katze.NOM.SG   flüchten.IPFV-PRS.3SG   vor Hund.NOM.SG
         'Dann flüchtet die Katze vor *der Hund.'
     e. Potom   kisa   uvidela   *ptenčiki.
         Dann   Mieze.NOM.SG   sehen.PFV-PRS-F.SG   Vögelchen.NOM.PL
         'Dann hat die Mieze die Vögelchen gesehen.'
     f. A   sobaka   *na košku   begaet.
         Und   Hund.NOM.SG   auf Katze.ACC.SG laufen.IPFV-PRS.3SG
         'Und der Hund läuft *auf die Katze.'

Auch in den Daten von 2016 haben wir einige Beispiele gefunden, *die Innovationen der Kindersprache* darstellen können:

(6) *Kljuvnu-l-a   za xvost.
      Schnabeln.PFV-PST.F.SG an Schwanz.ACC.SG
      '(Sie) hat nach dem Schwanz geschnabelt.'

Ein Kind benutzte auch Verb-Interjektionsformen statt Verben:

(7) a. I   potom   lisa   *xop.
         Und   danach Fuchs.NOM.SG hopp
         'Und danach *hopp der Fuchs.'

---

4   Statt: *Skazke* 'Märchen.DAT.SG' *konec* 'Ende.NOM.SG'. 'Das Märchen ist zu Ende.'

b. I   lisa   za nogu   kozlika
Und   Fuchs.NOM.SG an Bein.ACC.SG Ziegenböckchen.ACC.SG
*cap.
 schnapp
'Und der Fuchs *schnapp nach dem Bein des Ziegenböckchens.'

Die Kinder benutzen manchmal Mischsprache. Zum Beispiel haben zwei Kinder das Nomen *bejbiki* 'Babys' verwendet:

(8) a. Koza mama   ona   pomog-l-a   eë
         Ziege Mama.NOM.SG sie.NOM.SG helfen.PFV-PST-F.SG ihr
         *bejbiku.
          Baby.DAT.SG
         'Die Ziegen-Mama, sie hat ihrem Baby geholfen.'
     b. Tam oveči   dve *bejbiki   i   odna mama.
         Dort Schäfchen zwei Baby.NOM.PL und eine Mama.NOM.SG
         'Dort sind Schäfchen zwei Babys und eine Mama.'

Oder sie integrieren deutsche Substantive in den russischen Text:

(9) a. I   tam   byl   *Fuchs.
        Und   dort   sein.PST-M.SG   Fuchs.NOM.SG
        'Und dort war ein Fuchs.'
    b. Tam *Nest   est'   na dereve.
        Dort   Nest.NOM.SG sein.PRS.3SG auf Baum.LOC.SG
        'Dort gibt es ein Nest auf dem Baum.'

In diesem Beispiel benutzt ein Proband deutsche Wörter in einem russischen Satz, was als Borrowing bezeichnet werden kann:

(10) I   ona   fast sie war.
      Und   sie.NOM.SG
      ' Und sie - fast sie war.'

5.3 Datenanalyse: Russisch – Deutsch

Der Vergleich der lexikalischen Fähigkeiten der Kinder in beiden Sprachen lässt keine eindeutige Aussage dahingehend zu, dass Deutsch in der vorliegenden Probe die dominante Sprache darstellt. Einerseits zeigte sich bei der Betrachtung der gesamten Anzahl der Tokens sowohl im Jahre 2016, als auch im Jahre 2017 eine höhere Anzahl für das Deutsche (2017: 401; 2016: 794) als für das Russische

(2017: 248; 2016: 448). Dabei ist die Anzahl der Tokens in beiden Sprachen im Jahr 2016 in etwa doppelt so hoch wie im Jahr 2017, was vermutlich mit der Erhebungsform in Zusammenhang steht: 2016 wurden die Texte von den Kindern diktiert, 2017 selbst geschrieben.

Wenn man die Verteilung von Verben und anderen Wortarten betrachtet, sehen die Ergebnisse aus beiden Stichproben in beiden Sprachen fast gleich aus, wobei man sieht, dass die Probanden in den russischen Texten (2017: 20 %; 2016: 21 %) im Vergleich zu den deutschen Texten (2017: 16 %; 2016: 17 %) mehr Verben als andere Wortarten gebrauchen (vgl. Abb. 10). Dabei ist es interessant, dass der Anteil von der Erhebungsform relativ unabhängig zu sein scheint. Die Relation der Anzahl der Tokens und der Anzahl der Verben lässt darauf schließen, dass – eingedenk der Tatsache, dass die bilingualen Kinder auf Deutsch deutlich längere Texte produzieren, was wiederum auf eine reichere lexikalische Ausdrucksfähigkeit hinweist – der Anteil der Verben in den Stichproben aus beiden Sprachen fast gleich ist. Dieses Ergebnis bestätigt das Ergebnis von Klassert (2011: 160), das zeigt, dass im Vergleich zu Nomen, bei denen im Grundschulalter ein Sprachabbau in der Herkunftssprache beobachtbar ist, das Verb relativ stabil bleibt.

Anders sieht die Situation beim Vergleich der Anzahl von Verben, Hapax Legomina und präfigierten Verben in den russischen und deutschen Texten aus: In beiden Sprachen sind die jeweiligen Zahlen für die mündlich erhobenen Texte aus dem Jahr 2016 in etwa doppelt so hoch wie jene, der schriftlich erhobenen aus dem Folgejahr (vgl. Abb. 11). Vergleicht man die zwei Sprachen miteinander, so ist die absolute Anzahl der Verben (russisch 66; deutsch 68) sowie Hapax Legomina (Russisch: 58; Deutsch: 62) in deutschen und russischen Texten aus der Erhebungsrunde 2017 fast gleich. Im Gegensatz dazu ist die absolute Anzahl von Verben (Russisch 111; Deutsch 155) im Jahre 2016 in den deutschen Texten erheblich größer, wobei die Anzahl von Hapax Legomina (Russisch 87; Deutsch 91) einen solchen Unterschied nicht aufweist. Leider kann anhand dieser Daten nicht geschlossen werden, ob sich in beiden Sprachen bei den Verben innerhalb dieses Untersuchungszeitraumes von zwei Jahren, z. B. durch die längere Kontaktzeit mit L2 beeinflusst, ein Fortschreiten in der Entwicklung beobachten lässt, da die Daten in zwei unterschiedlichen Erhebungssettings (schriftlich vs. mündlich) erhoben wurden. Anschließend an die Ergebnisse der Studie von Klassert (2011) vermuten wir, dass die Verben in L1 ein geringes Fortschreiten und in L2 ein sichtbares Fortschreiten zeigen würden, obwohl die Anzahl der Verben sowie Hapax Legomina in beiden Sprachen gesunken ist, was wahrscheinlich damit zu erklären ist, dass die Probanden erst am Anfang des Schriftspracherwerbs stehen.
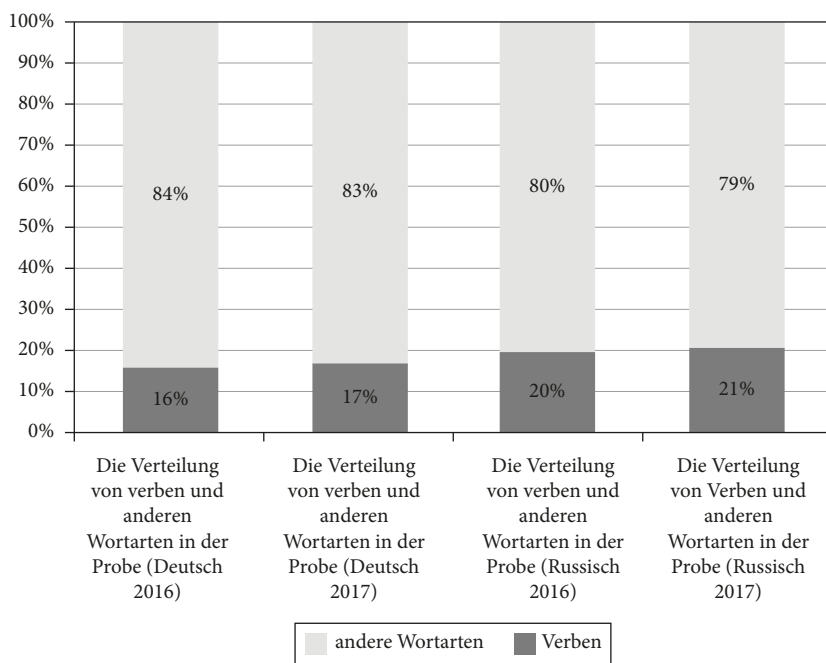
Abb. 9: Verteilung von Verben und anderen Wortarten im deutschsprachigen und russischsprachigen Datenkorpus 2016–2017

Wendet man die Zeitformenanalyse an, so kommt man beim Vergleich der Verteilung der Zeitformen zu folgenden Ergebnissen: In beiden Erhebungsrunden werden die Verben von bilingualen Kindern in russischsprachigen Texten vorwiegend im perfektiven Präteritum und in den deutschen Texten im Präsens gebraucht (vgl. Abb. 3 und Abb. 7). Auffallend bei diesen Ergebnissen ist, dass die Verteilung der Zeitformen in den russischen geschriebenen Erzählungen unseres Samples nicht mit der Verteilung der Zeitformen bilingualer Kinder, sondern mit jener russisch monolingualer Kinder aus der Studie von Anstatt (2008) korrespondiert. Sie hatte festgestellt, dass monolinguale russische Kinder zu 71 % auf das perfektive Präteritum als primäre Zeitform zurückgreifen. In unserer Stichprobe tun dies 68 %. Die Ergebnisse aus dem Jahr 2016 (diktierte Erzählungen) nähern sich jedoch den Ergebnissen von bilingualen Kindern bei Anstatt (2008): Nur 47 % der Fälle verwenden bei ihr das perfektive Präteritum als Hauptform (vgl. 54 % in der vorliegenden Studie). Betrachtet
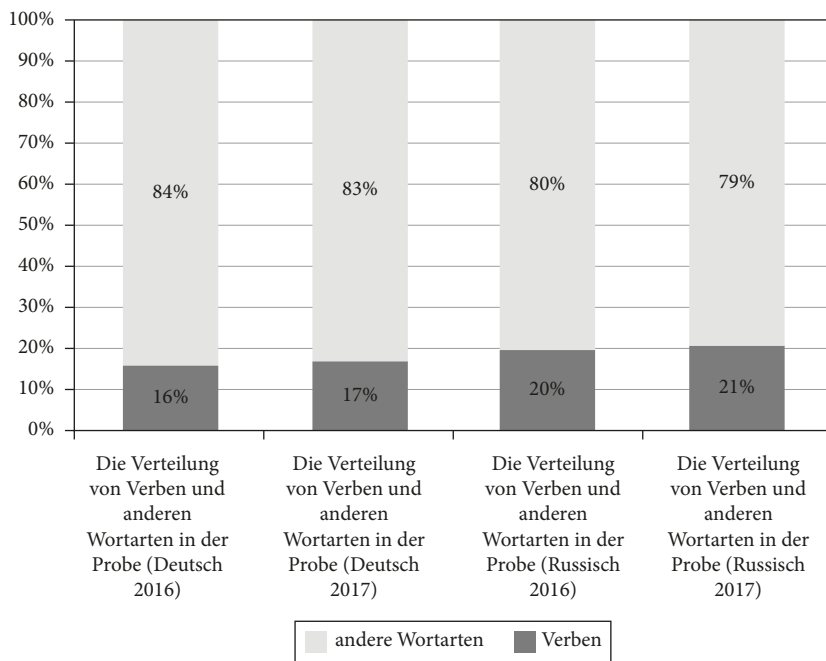
Abb. 10: Vergleich des Verbgebrauchs im deutschsprachigen und russischsprachigen Datenkorpus 2016–2017

man die Ergebnisse in Bezug auf Deutsch bei bilingualen Kindern, so stimmen die Befunde mit den Befunden von Anstatt (2008) überein: 60 % aller Fälle verwenden in ihrem Sample das Präsens als Hauptform (vgl. 52 % und 50 % in der vorliegenden Studie).

Vergleicht man die Verteilung des Zeitformengebrauchs pro Proband im deutschsprachigen und russischsprachigen Datenkorpus 2016 und 2017, erkennt man in Abb. 9 in Bezug auf die russischen Texte die gleiche Tendenz wie in Abb. 7: Abgesehen von einem haben alle Probanden das Präteritum als Haupterzählform ausgewählt, dabei haben im Jahr 2016 fünf von acht und im Jahr 2017 alle das perfektive Präteritum gegenüber dem imperfektiven Präteritum bevorzugt. Diese Ergebnisse entsprechen wieder den Ergebnissen zu monolingualen Kindern bei Anstatt (2008) und nicht jenen zu bilingualen Kindern. Bei der Betrachtung der deutschsprachigen Texte zeigt sich größere Varianz, obwohl man auf Abb. 3 eine eindeutige Dominanz des Gebrauchs der

Verben im Präsens sieht. So haben im Jahr 2016 fünf Kinder das Perfekt und drei Kinder das Präsens, im Jahr 2017 vier Kinder das Perfekt, drei Kinder das Präsens und ein Kind das Präteritum als Hauptform des Erzählens gewählt, was wiederum den Ergebnissen von bilingualen Kindern bei Anstatt (2008) widerspricht (74 % das Präsens, 24 % der Probanden das Perfekt). Die möglichen Ursachen für die genannten Unterschiede werden dann in der Diskussion beschrieben.

Was Interferenz anbelangt, so wurden in beiden Texten im Bereich des Verbs weder in den russischsprachigen, noch in den deutschsprachigen Texten Auffälligkeiten bemerkt, mit der Ausnahme vom Fall "fressen". Das trifft allerdings nicht auf andere Wortarten in den russischsprachigen Texten zu, in den bisweilen solche Effekte vorkommen. Zudem ist zu beachten, dass die schriftlichen russischsprachigen Texte sehr oft mit lateinischen Buchstaben verfasst wurden, da die Kinder die kyrillische Schrift nicht oder nicht ausreichend beherrschen. Entsprechende detaillierte Analysen stehen noch aus.

6 Diskussion und Fazit

In dieser Studie haben wir das Verbsystem von sequenziell bilingualen Kindern auf drei Ebenen (auf der lexikalischen und morphosyntaktischen sowie in Bezug auf Bilingualismus bedingte Interferenzerscheinungen) untersucht und sind zu den folgenden Ergebnissen gekommen:

Auf der lexikalischen Ebene zeigen die bilingualen Kinder insgesamt bessere lexikalische Fähigkeiten in L2 Deutsch als in L2 Russisch, wobei sich diese Differenz bei den Verben im Vergleich zu anderen Wortarten kaum manifestiert, was die These stützt, dass Verben bei dieser Probandengruppe im Vergleich zu anderen Wortarten, z. B. Nomen (vgl. Klassert 2011), relativ stabil bleiben. Leider konnten wir anhand unserer Daten keine Aussage darüber treffen, ob in beiden Sprachen im Beobachtungszeitraum von zwei Jahren, z. B. durch die längere Kontaktzeit mit L2, bei den Verben ein Fortschreiten in der Entwicklung stattfindet, da die Daten unterschiedlich erhoben wurden: 2016 sollten die Kinder die Geschichten diktieren, während sie sie 2017 selbst aufschreiben sollten. Man könnte allerdings vermuten, dass eben die deutsche Sprache zur dominanten Sprache wird, da die Probanden schon seit dem dritten Lebensjahr KITAs besucht haben, in denen sie einen unmittelbaren Kontakt mit L2 haben.

Auf der Ebene der Interferenzerscheinungen wurden auch Anzeichen bemerkt, jedoch nicht in Bezug auf das Verb, die dafür sprechen, dass für die Probanden eben die deutsche Sprache die dominante Sprache auf dieser Etappe des Spracherwerbs ist. Im Bereich der allgemeinen Lexik wurden in den

russischen, nicht aber in den deutschen Texten Beispiele für Interferenz offenbar, was als ein Anzeichen dafür gesehen werden kann, dass die russische Sprache nicht dominant ist.

Allerdings zeigte sich in den deutschen Texten eine interessante Tendenz bezüglich des Gebrauchs der Verben *essen/fressen* und deren präfigierten Formen in Bezug auf die Handlung von Tieren. Während sich im Russischen das Äquivalent für das Verb *essen* sowohl auf eine Person als auch auf ein Tier beziehen kann, wird im Deutschen bei Tieren primär *fressen* gebraucht, was in der Stichprobe nicht der Fall war. Man könnte das mit dem Einfluss von L1 auf L2 erklären. Hierfür bräuchte man jedoch eine größere Stichprobe und einen Vergleich mit monolingual deutschsprachigen Kindern der gleichen Altersgruppe.

Auf der morphosyntaktischen Ebene haben wir die Verben in Bezug auf die temporale Struktur der Geschichten untersucht. Dabei haben wir uns auf die Studie von Anstatt (2008) gestützt. Die Ergebnisse der Analyse von Tempus-Aspekt-Wahl in Narrativen von bilingualen Kindern im Alter von drei bis neun Jahren von Anstatt (2008) haben gezeigt, dass sich das bilinguale Erzählmodell in der Tempus-Aspekt-Wahl in der nicht dominanten Sprache (Russisch) deutlich an das der dominanten Sprache (Deutsch) annähert. Die Hälfte der Probanden wählte in ihren Erzählungen in L1 die typische Zeitform der L2, das (historische) Präsens, einige übertrugen das russischsprachige Model (Präteritum) ins Deutsche und verwendeten das Perfekt und nur drei von 30 Probanden wählten die für die Sprache typische Erzählform (Deutsch: Präsens, Russisch: Präteritum, vgl. Anstatt 2017).

Die Ergebnisse unserer Studie unterschieden sich jedoch von den gerade genannten. Die Mehrheit der Probanden wählte das Präteritum, die typische Erzählform für das Russische, als Haupterzählform in ihren russischsprachigen Texten sowohl im mündlichen als auch im schriftlichen Erhebungssetting. Dabei wurde das perfektive Präteritum dem imperfektiven Präteritum vorgezogen. Wenn in der Studie von Anstatt (2017) nur einige Kinder beim Erzählen das russischsprachige Model (Präteritum) in die deutschsprachigen Texte übernommen und das Perfekt gebraucht haben, machten das mehr als die Hälfte der Probanden unserer Studien in beiden Erhebungsrunden. Nur drei von acht Kindern benutzten die für die deutsche Sprache typische Form (Präsens). Man kann somit sagen, dass unsere Studie eine andere Tendenz als die von Anstatt (2008) zeigt: Die L1 (Russisch) bleibt noch in diesem Bereich des Verbgebrauchs absolut dominant und hat einen starken Einfluss auf die temporale Struktur der L2 (Deutsch).

Diese klaren Unterschiede zwischen beiden Sprachen können auf eine unterschiedliche Struktur der Studienpopulation zurückgeführt werden. So

haben die Probanden verschiedene Ausgangsvoraussetzungen in Bezug auf den ethnolinguistischen Aspekt, der eine große Rolle im bilingualen Spracherwerb spielt (vgl. Tchirscheva 2012, S. 65): In Anstatts (2008) Studie haben neun von 30 Probanden deutschsprachige Väter (bi-ethnischer Bilingualismus), in der vorliegenden Studie hingegen haben beide Elternteile der Probanden Russisch als L1 (mono-ethnischer Bilingualismus). Diese Tatsache beeinflusst auch die Kontaktzeit mit L2. Die Kinder, deren Väter Deutsch als L1 haben, haben den Kontakt zur deutschen Sprache von Geburt an und wachsen simultan-bilingual auf, im Gegensatz dazu wachsen die Kinder mit russischsprachigen Eltern sequentiell-bilingual auf, da sie später in den Kontakt mit L2 treten. Das Alter der Kinder spielt bei der Stichprobe auch eine große Rolle: Je länger die Kinder im Kontakt mit L2 sind, desto größer ist der Einfluss dieser Sprache auf L1. So waren die Probanden aus unserer Stichprobe zwischen sechs und sieben Jahre alt, während die Altersspanne aus der Probe von Anstatt (2008) viel größer war – von 3,1 bis 9,6 Jahre, darunter waren nur sechs Probanden der Altersgruppe aus unserer Stichprobe. Auch die Erhebungsform (schriftlich vs. mündlich) könnte einen Einfluss auf die Wahl der Haupterzählform haben. Die oben aufgestellten Hypothesen gilt es in Folgestudien zu überprüfen.

Zusammenfassend kann man sagen, dass bilinguale Kinder nicht die gleichen Gebrauchsmuster von Verben und Verbalkategorien in der Sprachproduktion in beiden Sprachen zeigen. Sehr oft kommt es zu Interferenz. So nähern sich die deutschsprachigen Texte dem russischen Tempusmodell an. Auf diese Weise entwickelt sich das Verbsystem in Bezug auf die betrachteten Kategorien bei bilingualen Kinder verschieden, aber nicht getrennt, da beide Sprachen einen Einfluss aufeinander haben.

Literatur

Anstatt, Tanja: Aspect and tense in storytelling by Russian, German and bilingual children. *Russian Linguistics* 32/1, 2008, S. 1–26.

Anstatt, Tanja: Entwicklung des Russischen in Deutschland: Der Verbalaspekt im Russischen als Herkunftssprache. In: Witzlack-Makarevich, Kai/Wulff, Nadja (Hrsg.): *Handbuch des Russischen in Deutschland*: *Migration – Mehrsprachigkeit – Spracherwerb*. Berlin: Frank&Timme. 2017. S. 303–321.

Anstatt, Tanja/Dieser, Elena: Sprachmischung und Sprachtrennung bei zweisprachigen Kindern (am Beispiel des russisch-deutschen Spracherwerbs). In: Anstatt, Tanja (Hrsg.): *Mehrsprachigkeit bei Kindern und Erwachsenen. Erwerb, Formen, Förderung*. Tübingen: Attempto. 2007. S. 139–162.

Belliveau, Corinna: *Simultaner bilingualer Spracherwerb unter entwicklungs- und kognitionspsychologischen Aspekten*. Aachen: Shaker Verlag. 2002.

Bericht des Berliner Interdisziplinären Verbunds für Mehrsprachigkeit (BIVEM). Oktober 2011 - Juni 2014, zugegriffen am 12.12.2018, auf http://www.zas.gwz-berlin.de/fileadmin/projekte/bivem/bivembericht-kurz.pdf.

David, Annabelle/Wei, Li: The composition of the bilingual lexicon. In: Cohen, James/ McAlister, Kara T./Rolstad, Kellie/MacSwan, Jeff (Hrsg.): *ISB4: Proceedings of the 4th International Symposium on Bilingualism*. Somerville: Cascadilla Press. 2005, S. 594–607.

Deuchar, Margaret/Quay, Suzanne: Language choice in the earliest utterances: A case study with methodological implications. *Journal of child language* 26/2, 1999, S. 461–475.

Gagarina, Natalia: Verbs of motion in Russian: An acquisitional perspective. *Slavic and East European Journal* 53, 2009, S. 351–370.

Gagarina, Natalia: Acquisition and loss of L1 in a Russian-German bilingual child: A case study. In: Cejtlin, Stella (Hrsg.): *Monolingual and bilingual path to language*. Moskau: Jazyki slavjanskoj kul'tury. 2011. S. 137–163.

Gagarina, Natalia: Sprachdiagnostik in der Erstsprache mehrsprachiger Kinder (am Beispiel des Russischen). *Sprache. Stimme. Gehör* 37, 2013, S. 196–200.

Gagarina, Natalia: Monolingualer und bilingualer Erstspracherwerb des Russischen: ein Überblick. In: Witzlack-Makarevich, Kai/Wulff, Nadja (Hrsg.): *Handbuch des Russischen in Deutschland: Migration – Mehrsprachigkeit – Spracherwerb*. Berlin: Frank&Timme. 2017. S. 393–410.

Gagarina, Natalia/Klop, Daleen/Kunnari, Sari/Tantele, Koula/Välimaa, Taina/Balčiūnienė, Ingrida/Bohnacker, Ute/Walters, Joel: *MAIN: Multilingual Assessment Instrument for Narratives*. Berlin: ZAS. 2012.

Gagarina, Natalia/Klop, Daleen/Kunnari, Sari/Tantele, Koula/Välimaa, Taina/Balčiūnienė, Ingrida/Bohnacker, Ute/Walters, Joel: Assessment of Narrative Abilities in Bilingual Children. In: Armon-Lotem, Sharon/de Jong, Jan/Meir, Natalia (Hrsg.): *Assessing Multilingual Children: Disentangling bilingualism from language impairment*. Bristol: Multilingual Matters. 2015. S. 243–274.

Gagarina, Natalia/Posse, Dorothea/Düsterhöft, Stefanie/Topaj, Nathalie/Acikgöz, Duygu: Sprachförderung bei Mehrsprachigkeit: Erste Ergebnisse der BIVEM-Studie zur Wirksamkeit von Sprachförderung bei jüngeren mehrsprachigen Kindern: Eine Studie des Berliner Interdisziplinären Verbundes für Mehrsprachigkeit (BIVEM). *Spektrum Patholinguistik* 7, 2014, S. 139–148.

Hepsöyler, Ender/Liebe-Harkort, Klaus: *Muttersprache und Zweitsprache: türkische Schulanfängerinnen und Schulanfänger in der Migration. Ein Vergleich*. Frankfurt/Main: Peter Lang. 1991.

Iluz-Cohen, Peri/Walters, Joel: Telling stories in two languages: Narratives of bilingual preschool children with typical and impaired language. *Bilingualism: Language and Cognition* 15(1), 2012, S. 58–74.

Jia, Gisela/Kohnert, Kathryn J./Collado, Julissa/Aquino-Garcia, Francia: Action naming in Spanish and English by sequential bilingual children and adolescents. *Journal of Speech, Language, and Hearing Research* 49(3), 2006, S. 588–602.

Karasu, Ibrahim: *Bilinguale Wortschatzentwicklung türkischer Migrantenkinder vom Vor-bis ins Grundschulalter in der Bundesrepublik Deutschland*. Frankfurt am Main: Peter Lang. 1995.

Klassert, Annegret: *Lexikalische Fähigkeiten bilingualer Kinder mit Migrationshintergrund. Eine Studie zum Benennen von Nomen und Verben im Russischen und Deutschen*. Marburg. 2011.

Kohnert, Kathryn J./Bates, Elizabeth/Hernandez, Arturo E.: Balancing bilinguals: Lexical-semantic production and cognitive processing in children learning Spanish and English. *Journal of Speech, Language, and Hearing Research* 42(6), 1400–1413. 1999.

Leist-Villis, Anja. o.J.: *Information – Fortbildung – Netzwerk zur frühkindlichen Zweisprachigkeit,* zugegriffen am 12.12.2018, auf http://www.zweisprachigkeit.net/zweisprachigkeit_ist.htm.

Schmid, Monika: *First language attrition, use and maintenance: The case of German Jews in Anglophone countries*. Amsterdam: John Benjamins Publishing 2002.

Sheng, Li/McGregor, Karla K./Marian, Viorica: Lexical-semantic organization in bilingual children: Evidence from a repeated word association task. *Journal of Speech, Language and Hearing Research* 49(3), 2006, S. 572–587.

Tchirsheva, Galina N.: *Childhood bilingualism: the simultaneous acquisition of two languages*. St. Petersburg: Zlatoust. 2012.

Van der Linden, Elisabeth: Non-selective access and activation in child bilingualism: the Lexicon. In: Döpke, Susanne (Hrsg.): *Cross-linguistic structures in simultaneous Bilingualism*. Amsterdam: John Benjamins 2000. S. 37–57.

Tamás Tölgyesi

Hungarismen im Gemeindeutschen, österreichischen Deutsch, ostösterreichischen Dialekt und im Slawischen

Abstract: In this study, I examine only those Hungarian loanwords that exist in different varieties of the German language, as well as in several Slavic languages. Hungarian loanwords can be divided into three groups. The first group contains Hungarisms, which are known throughout the German-speaking countries and in most Slavic countries (including Russia), e.g. Ger. *Gulasch*, Cz. *guláš*, Slk. *guláš*, *guľáš*, Pol. *gulasz*, Sl. *golaž*, B/K/S *gulaš*, Rus. *гуляш* 'goulash' < Hun. *gulyás (hús)* 'meat of cattle herders', to Hun. *gulya* 'cattle herd'. The second group includes such expressions that occur in the languages of the peoples of the former Habsburg Monarchy, e.g. Austrian Ger. *Palatschinke*, Cz./Sl./B/K/S *palačinka*, Slk. *palacinka*, Pol. *palaczinka* 'pancake' < Hun. *palacsinta* 'pancake' < Rum. *plăcintă* '(apple) pie'< Lat. *placenta* 'cake', 'uterine cake'. In the third group there are the loanwords from the Hungarian, which existed or are still used mainly in the neighbouring areas of present-day Hungary (in Burgenland, Serbia and Slovakia, where also Hungarian minority lives), e.g. East Austrian dial. *Hotter*, Slk. *chotár*, B/K/S *hatar*, *atar* 'county line' < Hun. *határ* 'border'.

Keywords: loanwords, Hungarian, German, Slavic, historical sociolinguistics

1 Einleitung

Der vorliegende Vortrag erwuchs aus einer von mir gehaltenen Vorlesung aus dem Sommersemester 2017 mit dem Titel „Austriazismen – gemeinsames Lehngut im Deutschen, Slawischen und Ungarischen", die nicht nur SlawistInnen und FinnougristInnen, sondern auch Studierende des Faches „Austrian Studies" besuchten. In diesem Rahmen wurden auch ungarische Lehnwörter im Deutschen und Slawischen behandelt. Im Gegenteil zu Germanismen, die im Ungarischen, Tschechischen und anderen slawischen Sprachen sehr häufig vorkommen, gibt es nur eine geringe Anzahl von Hungarismen, die sowohl im Deutschen, als auch in den Slawinen gebräuchlich sind.

Ungarische Lehnwörter können in drei Gruppen eingeteilt werden. Die erste Gruppe bilden Hungarismen, die im ganzen deutschsprachigen Raum und in den meisten slawischen Ländern (also auch in Russland) bekannt sind. Zur zweiten Gruppe zählen solche Ausdrücke, die in den Sprachen der Völker der

ehemaligen Habsburgermonarchie vorkommen.[1] In der dritten Gruppe befinden sich Lehnwörter aus dem Ungarischen, die vor allem in den benachbarten Gebieten des heutigen Ungarns (im Burgenland, Serbien und in der Slowakei, wo auch eine ungarische Minderheit lebt, siehe dazu Tóth 2019) existierten oder bis heute verwendet werden.

2 Hungarismen im Gemeindeutschen und Slawischen

Dt. *Attila*, tsch./slk./sln./bks. *atila* 'Husarenrock, Waffenrock; (nicht als Uniform) mit Schnüren besetzte Jacke' sind aus ung. *atilla* 'mit Schnüren besetzte ungarische Jacke' entlehnt. Der Name des Hunnenkönigs Attila, Atilla wurde appellativiert und für das gegebene Kleidungsstück verwendet. Früher wurde dieses Proprium mit Doppel-L, heute mit Doppel-T geschrieben, aber die ursprüngliche Aussprache [ɒtilːɒ] wurde beibehalten.

Dt. *Csárda*, slk. čárda (Králik 2016, S. 100), bks. *čarda* 'Heideschenke' stammen von ung. *csárda*. In der Csárda wird oft Csárdás gespielt und getanzt. Dt. *Csardas*, *Csárdás*, tsch./slk. *čardáš*, pol. *czardasz*, sln./bks. *čardaš*, rus. чардаш 'ungarische Volksmusik und Nationaltanz' sind aus ung. *csárdás* übernommen.

Dt. *Gulasch*, tsch. *guláš*, slk. *guláš*, *guľáš*, pol. *gulasz*, sln. *golaž*, bks. *gulaš*, rus. гуляш sind im 19. Jh. aus ung. *gulyás(hús)* 'Fleisch der Rinderhirten', zu ung. *gulya* 'Rinderherde' entlehnt. Ung. *gulyás* bezeichnet sowohl den Rinderhirten[2], als auch eine Art dicker Suppe *gulyásleves* 'Gulaschsuppe'. Dem deutschen, tschechischen und slowakischen Gulasch ähnelt am meisten das sog. ungarische Pörkölt, welches die Ungarn mit Nockerln, Teigwaren oder Kartoffeln essen, keineswegs mit Knödeln wie die Tschechen. Das sog. Szegediner und Debracziner Gulasch[3] (tsch. *segedínský a debrecínský guláš*) kennen nur diejenigen Bewohner der ungarischen Großstädte Szeged und Debrecen, die diese Spezialitäten während ihres Aufenthaltes im Ausland gekostet haben (Tölgyesi 2009a, S. 19, 2009b, S. 146). Für weitere Gulaschvarianten, wie z.B. Pressburger Gulasch, Znaimer Gulasch, Triester Gulasch, siehe Pohl (2007, S. 72–74, 2017, S. 230–231). Ukr. бограч 'Gulaschsuppe' und sln. *bograč* 'Kesselgulasch' sind über ung.

---

1   Zum mitteleuropäischen Sprachbund vgl. Kurzová 1996, Pilarsky 2001, Newerkla 2002, Bláha 2015, Januška 2017.

2   Eine ähnliche Bildung ist dt. *Csikos*, *Csikós*, bks. *čikoš* aus ung. *csikós* 'Pferdehirt', zu ung. *csikó* 'junges Pferd, Fohlen'.

3   Das Debrecziner Gulasch wurde nicht nach seinem Ursprungsort, sondern nach seinem wichtigsten Bestandteil – Scheiben von Debrecziner Wurst – benannt.

bogrács 'Kessel' und bks. *bakrač* aus tür. *bakraç* 'Kupfernapf' übernommen (zu näheren Einzelheiten vgl. Kocsis 2018).

Dt. *Husar*, tsch. *husar*, slk. *husár*, pol. *husarz, huzar*, bks. *husar*, rus. *гусар* 'Angehöriger der leichten Reiterei in ungarischer Nationaltracht; leicht bewaffneter Reiter' gehen auf ung. *huszár* zurück. Das ungarische Wort ist über die älteren serbischen Formen *husar, gusar* '(See)räuber' aus it. *corsaro* 'Korsar' entlehnt.

Dt. *Kutsche*, tsch. *kočár*, slk. *koč, kočiar*, pol. *kocz*, sln. *kočija*, bks. *kočije* (Plur.), erst später auch *kočija* (Sing.) 'von Pferden gezogener, meist geschlossener Wagen zur Beförderung von Personen' stammen von ung. *kocsi*, zu *kocsi szekér* 'Wagen aus dem Ort Kocs'. Ung. *kocsi* 'Wagen, Kutsche', bezeugt seit Ende des 15. Jh.; substantiviertes Adjektiv, gebildet mit dem Suffix *-i* aus dem Ortsnamen Kocs, Dorf im Komitat Komárom, dessen Einwohner mit ihren schnellen Fuhrwerken im Verkehr zwischen Wien und Ofen (d.h. Buda) eine wichtige Rolle spielten. Das ung. Wort drang in fast alle europäischen Sprachen ein (Hadrovics 1985, S. 315–316). Im Deutschen werden scherzhaft auch größere und alte Autos als Kutsche bezeichnet. Demgegenüber wird *kocsi* in der ungarischen Umgangssprache auch für neue Autos gebraucht. In der ungarischen Standardsprache funktioniert *kocsi* u.a. als Teil der Komposita *lovaskocsi* 'Pferdekutsche', *személygépkocsi* 'Personenkraftwagen', *tehergépkocsi* 'Lastkraftwagen'.

Dt. *Paprika*, tsch./slk. *paprika*, pol. *papryka* sind über ung. *paprika* aus bks. *paprika* (vgl. sln. *paprika*), zu bks. *papar* 'Pfeffer' entlehnt. Dt. *Paprikasch*, slk. *paprikaš*, pol. *paprykarz*[4], bks. *paprikaš* stammen von ung. *paprikás* '(Hühner) gulasch', eigentlich *paprikás (hús)* 'papriziertes (Fleisch)'.

Dt. *Puszta*, slk. pusta (Králik 2016, S. 484), bks. *pusta*, rus. *пуста* 'ungarische Steppe' stammen von ung. *puszta*, welches aus urslaw. Neutr. *pusto*, Fem. *pusta*, zu *pustъ* 'leer' übernommen ist.

Dt. *Tol(l)patsch*, slk. *talpaš* (Rocchi 2010, S. 55), bks. *talpač* (Hadrovics 1985, S. 491) 'Fußsoldat' sind aus ung. *talpas* '(breit)füßig', zu *talp* 'Sohle' entlehnt und bezeichnete früher die ungarischen Fußsoldaten, die statt Schuhen mit Schnüren befestigten Sohlen trugen (Kluge 2002, S. 919). Die heutige Bedeutung 'ungeschickter Mensch' entstand durch Volksetymologie aus den Wörtern *Tölpel* 'ungeschickter Mensch' und *patschen* '(gehen, laufen und dabei) ein klatschendes Geräusch hervorbringen'. Tsch./slk. *ťulpas*[5] 'ungeschickter

---

4 In Polen werden Fischkonserven unter dem Namen *paprykarz szczeciński* 'Stettiner Paprikasch' hergestellt.

5 Auf den Zusammenhang zwischen dt. *Tollpatsch* und tsch./slk. *ťulpas* hat mich der Vorstand des Lehrstuhls für Osteuropäische Studien, Dr. Marián Sloboda, aufmerksam

Mensch' ist wahrscheinlich über dt. *Tollpatsch* aus ung. *talpas* entlehnt (Rejzek 2001, S. 682–683). Die heutige Bezeichnung für Fußsoldaten ist ung. *gyalogos (katona)*.[6]

Dt. *Tschako* 'helmartige Kopfbedeckung beim Heer und Polizei', tsch. *čáka*[7] (Fem.), *čáko* (Neutr.), slk. *čakov*, pol. *czako*, sln. *čaka*, bks. *čaka*, *čako* sind aus ung. *csákó* 'Husarenhelm' übernommen.

3 Hungarismen im österreichischen Deutsch und im Slawischen

Ödt. *Fogosch*, *Fogasch* 'Zander' ist aus ung. *fogas* 'Zander', zu *fogas (hal)* 'gezahnter (Fisch)' entlehnt. Tsch. *candát* und pol. *sandacz* stammen von dt. *Zander* (Brückner 1974, S. 481), aber im tschechischen Fischerslang wird das ungarische Lehnwort *fogoš* gebraucht (Hubáček 2003, S. 60). Slk. *zubáč* 'Zander' wird aus dem Wortstamm *zub* 'Zahn' mit dem Suffix *-áč* gebildet. Ung. *fogas* mit der Bedeutung 'Kleiderhaken' wurde ins Slowakische als *fogaš* entlehnt.

Ödt. *Langosch*, pol. *langosz*, tsch., slk. und sln. *langoš* sind aus ung. *lángos*, zu *láng* 'Flamme' entlehnt. Früher wurde Langosch nicht in Öl, sondern nahe der Flamme gebacken. Langosch isst man in Ungarn mit Knoblauchsaft oder Sauerrahm und geriebenem Käse.

Ödt. *Palatschinke*, tsch./sln./bks. *palačinka*, slk. *palacinka* und pol. *palaczinka* sind über ung. *palacsinta* und rum. *plăcintă* [platʃinta] aus lat. *placenta* entlehnt (Zaic 2006, S. 608), das im Allgemeinen 'Kuchen', in der Medizin 'Mutterkuchen' bedeutet. Interessant ist, dass plăcintă in Rumänien kein Pfannkuchen, sondern ein Käse- oder Apfelkuchen ist (Tölgyesi 2017a, S. 273).

Ödt. *Schinakel* 'kleines Ruderboot' ist über ung. *csónak* 'Boot, Kahn' aus den slawischen Sprachen entlehnt, vgl. tsch. *člun*, slk. *čln*, pol. *czółno*, sln. *čoln*, rus. *челнок* < urslaw. *čьlnъ*.[8] Als Pluraletantum wird ödt. *Schinakel(n)* auch im Sinne von 'breite, ausgetretene Schuhe' verwendet. In der ungarischen Umgangssprache hat *csónak* auch die zusätzliche Bedeutung '(zu) große Schuhe'. Der österreichische Schriftsteller Franz Dungl schreibt in seinem Gedicht „Des Schinakl" (08.01.2013) über die Herkunft und Bedeutung des erwähnten Wortes:

---

gemacht, als ich im Rahmen eines CEEPUS-Stipendiums an der Karls-Universität Prag einen Vortrag zum Thema Hungarismen gehalten habe (27.03.2018).

6   Das deadjektivische Substantivum *gyalogos* selbst bedeutet auch 'Fußgänger'.

7   Tsch. *čáka* 'Tschako' und *čáka* 'Erwartung' sind Homonyme.

8   Im B/K/S wird statt dieses gemeinslawischen Wortes der Ausdruck *čamac* (zu dial. *čam* 'Tanne') gebraucht, der aus tür. *çam* 'Tanne' übernommen ist.

Schinakl kummt, hot man mia gsogt,
von csonak – nur, wenn ana frogt.
In Ungarn nennt man so den Kahn,
den ma vom Laund tuat aufezahn.
Ob Boot, ob Schiff a gaunz a klanes,
doch untergehn, jo, jo, des kann es.
Und, wenn i daunn no weita suach,
so kumm i a auf große Schuach.
Des san bei uns die großn Latschn,
mit de d'Leut umaraunda hatschn.
Du kaunnst, modern, sie Moonboots nenna,
Du wirst bestimmt den Ausdruck kenna.
Und diese Kunststoff-Stiefel sind,
nicht ledern, vielleicht von an Rind.
Es tragen auch die größten Lackeln,
die Schuach, de man nennt a Schinakeln.

4 Hungarismen im ostösterreichischen Dialekt und im Slawischen

Ostö. *Aldomasch*, tsch. *aldamaš*, *haldamaš*, slk. *oldomáš*, bks. *aldomaš*, *jaldomaš* (Hadrovics 1985, S. 113–114) sind aus ung. *áldomás* 'Kauftrunk; Segnen, Segnung', zu *áld* 'segnen' entlehnt. Im Wörterbuch der deutschen Winzersprache steht folgendes:

Aldomasch 1. Ausspruch, um anzuzeigen, dass die Lese zu Ende ist (Burgenland), 2.a. (Fest)mahlzeit zum Abschluss der Lese (NÖ), b. kl. familiäres Leseabschlussfest mit den Hilfskräften (Burgenland), c. Umtrunk mit Wein u. kl. Speisen zum Abschluss der Lese im Weinkeller, der sich im Weinberg befindet (Mór), 3. Freiwein zu bes. Gelegenheiten, Kauftrunk (Budapest), Etym.: entl. aus ung. *áldomás* 'Trinkspruch; Gelage, Festschmaus; Kauftrunk'.

Ostö. *Gate(hose)* 'lange, overallähnliche Unterhose (für Männer)', tsch. *gatě*, slk. *gate*, pol. *gacie*, sln. *gate*, bks. *gaće*, rus. *гачи* sind wahrscheinlich aus ung. *gatya* 'Unterhose' übernommen. Im österreichischen Deutsch sagt man bis heute *Untergate* 'Unterhose'. In der tschechischen, slowakischen und ungarischen Umgangssprache werden *gatě*, *gate* bzw. *gatya* expressiv für Hosen (auch z.B. für neue Jeans) gebraucht.

Ostö. *Hotter*, slk. *chotár*, bks. *hatar*, *atar* (Hadrovics 1985, S. 260) 'Gemeindegrenze' stammen von ung. *határ* 'Grenze'.

Ostö. *Maschekseite* 'entgegengesetzte Seite, Rückseite' ist ein hybrides Kompositum aus ung. *másik* 'andere' und dt. *Seite*.

Ostö. *Mulatschag, Mulatschak*, bglk. *mulačok* (Kinda-Berlakovich 2004, S. 148) 'ausgelassenes Fest' ist aus ung. *mulatság* 'Belustigung' entlehnt. Daneben existieren ödt. *mulatieren* und ung. *mulat* 'an einem Mulatschag teilnehmen, ausgiebig feiern'. Bglkr. *mulatovati* bedeutet 'zechen' (Vig 2007, S. 112).

Ostö. *Salasch*, bglkr. *salaš* 'Schweinestall', tsch./slk./bks. *salaš* 'Herberge', pol. *szałas,* † *szałasz,* † *sałasz*[9] (Krasnowolski – Niedźwiedzki 1920, S. 583) sind aus ung. *szállás* 'Unterkunft' übernommen. Es geht um einen sog. Karpatismus, d.h. ein Wort aus dem Alltagsleben der Walachen, die aus dem Fürstentum Walachei (heute Rumänien) stammende Schafhirten waren und Siedlungen in den Karpaten (Karpatenukraine, Polen, Slowakei, Tschechien: *Valašsko* 'Mährische Walachei') und auf der Balkanhalbinsel hatten.[10] Bei diesem Ausdruck sind unterschiedliche Bedeutungsveränderungen zu beobachten. Die Bedeutung des ostö. *Salasch* und bglkr. *salaš* hat sich mit der Zeit verschlechtert und wird heute im Sinne 'Schweinestall' verwendet (vgl. ÖW, Vig 2007, S. 113).[11] Demgegenüber kann ung. *szállás* auch ein luxuriöses Fünfsternenhotel bezeichnen. In diesem Fall ist es ein schönes Beispiel für Bedeutungsverbesserung.

In diesem Beitrag wollte ich nur diejenigen ungarischen Lehnwörter untersuchen, die sowohl in verschiedenen Varietäten der deutschen Sprache, als auch in mehreren Slawinen vorkommen. Darüber hinaus gibt es noch eine Reihe von Hungarismen, die in den slawischen Sprachen – vor allem im Slowakischen[12] und B/K/S – verwendet werden, wie z.B. slk./bks. *bosorka* 'Hexe' < ung. *boszorka* 'Hexe', tsch. dial./slk. *gazda* 'Landwirt, Besitzer', bks. *gazda* 'Chef, Hausherr, Vermieter' < ung. *gazda* 'Landwirt, Besitzer' < slaw. *gospoda*, slk.

---

9   Ung. *sz* wird als [s], *s* als [ʃ] ausgesprochen. Im Polnischen funktioniert es umgekehrt. Pol. *sz* steht für [ʃ], *s* für [s]. Die ältere polnische Form *sałasz* [sawaʃ] ähnelte der ungarischen Aussprache *szállás* [saːlːaːʃ]. Die heutige polnische Version *szałas* [ʃawas] folgt der ung. Schreibweise *szállás*.

10 Weitere Karpatismen sind z.B. tsch. *baganče*, slk. *baganča* 'Stiefel' < ung. *bakancs*, tsch./ slk. *fujara*, bks. *frula*, ung. *furulya* 'Schnabelflöte' (näheres siehe OKDA, Vašek 2001, S. 131). Pol. *fujara* hat neben 'Flöte' die pejorative Bedeutung 'Idiot, Flasche, unfähiger Mensch'.

11 Einen großen Dank möchte ich an meinen Kollegen Andreas Gellan (Universität Wien, Österreichische Akademie der Wissenschaften) richten, der meine Hypothese zur Verbreitung des Wortes *Sal(l)asch* im Burgenland durch eine gründliche Recherche im „Wörterbuch der bairischen Mundarten in Österreich" bestätigt hat.

12 Neben *čákov* und *guláš, guľáš* behandelt Newerkla (2011, S. 593–594) auch slk. *harc* 'Scharmützel', *korheľ* 'Säufer', dial. *lenča* 'Linse', *mincier* 'Schnellwaage', *parta* 'Jungfernkranz, Kopfputz', *pohár* 'Becher, Pokal', *želiar* 'Häusler' als mögliche Entlehnungen aus dem Ungarischen.

sersám 'Pferdegeschirr', 'Werkzeug', bks. *sersam*, *sersan* 'Pferdegeschirr' < ung. *szerszám* 'Werkzeug' (vgl. auch ung. *lószerszám* 'Pferdegeschirr').

In mehreren etymologischen Wörterbüchern und kontaktlinguistischen Publikationen werden Hungarismen falsch, d.h. mit orthographischen Fehlern und inadäquater Bedeutung, angegeben (siehe dazu Tölgyesi 2017b, 2018). Es ist geplant, in Zukunft mit Dr. Jiří Januška, einem Hungarologen an der Karls-Universität Prag, gemeinsam zu ungarischen Lehnwörtern im Tschechischen zu forschen.

Abkürzungen für Sprachen und Dialekte

bglkr. = burgenlandkroatisch, bks. = bosnisch/kroatisch/serbisch, dt. = deutsch, it. = italienisch, lat. = lateinisch, ödt. = österreichisch, ostö. = ostösterreichisch, pol. = polnisch, rus. = russisch, slaw. = slawisch, slk. = slowakisch, sln. = slowenisch, tsch. = tschechisch, tür. = türkisch, ukr. = ukrainisch, ung. = ungarisch, urslaw. = urslawisch

Literatur

Bláha, Ondřej: *Jazyky střední Evropy*. Olomouc: Univerzita Palackého. 2015.

Brückner, Aleksander: *Słownik etymologiczny języka polskiego*. Warszawa: Wiedza Powszechna. 1974.

Hadrovics, László: *Ungarische Elemente im Serbokroatischen*. Budapest: Akadémiai Kiadó. 1985.

Hubáček, Jaroslav: Výběrový slovník českých slangů (= Spisy Filozofické fakulty Ostravské univerzity 146). Ostrava: Ostravská univerzita. 2003.

Januška, Jiří: *Porovnávání středoevropských jazyků: za horizont strukturních rysů a lexikálních přejímek*. Praha: Univerzita Karlova. 2017.

Kinda-Berlakovich, Andrea Zorka: Deutsche Interferenzen im Burgenlandkroatischen. In: *Zagreber Germanistische Beiträge, Beiheft* 7. 2004. S. 141–154.

Kluge, Friedrich: *Etymologisches Wörterbuch der deutschen Sprache*. Berlin: De Gruyter. 2002.

Kocsis, Mihály: A szláv nazálisok és a huszár. Magyar Nyelv 114, 2018, S. 68–77.

Králik, Ľubor: *Stručný etymologický slovník slovenčiny*. Bratislava: Veda. 2016.

Krasnowolski, Antoni/Niedźwiedzki, Władysław: *Słownik staropolski*. Warszawa: Wydawnictwo M. Arcta. 1920.

Kurzová, Helena: Mitteleuropa als Sprachareal. *Germanistica Pragensia* 13, 1996, S. 57–73.

Newerkla, Stefan Michael: Sprachliche Konvergenzprozesse in Mitteleuropa. In: Pospíšil, Ivo (Hrsg.): *Crossroads of Cultures: Central Europe/Perekrestki kuľtury: Srednjaja Evropa/Křižovatky kultury: Střední Evropa* (= Litteraria Humanitas XI). Brno: Masarykova univerzita. 2002. S. 211–236.

Newerkla, Stefan Michael: *Sprachkontakte Deutsch – Tschechisch – Slowakisch. Wörterbuch der deutschen Lehnwörter im Tschechischen und Slowakischen: historische Entwicklung, Beleglage, bisherige und neue Deutungen* (= Schriften über Sprachen und Texte 7). Frankfurt am Main: Peter Lang. 2011.

ÖW = Ostarrichi: Österreichisches Wörterbuch, zugegriffen am 01.03.2019, auf http://www.ostarrichi.org/.

OKDA = *Обшчекарпатский диалектологический атлас*. 1987–2003.

Pilarský, Jiří: *Donausprachbund. Das arealistische Profil einer Sprachlandschaft*. Debrecen: Universität Debrecen. 2001.

Pohl, Heinz-Dieter: *Die österreichische Küchensprache. Ein Lexikon der typisch österreichischen kulinarischen Besonderheiten (mit sprachwissenschaftlichen Erläuterungen)* (= Studia Interdisciplinaria Ænipontana 11). Wien: Praesens Verlag. 2007.

Pohl, Heinz-Dieter: Die österreichische Küchensprache im Lichte von Sprachgeographie und Sprachkontakt. In: Szczęk, Joanna/Kałasznik, Marcelina (Hrsg.): *Intra- und interlinguale Zugänge zum kulinarischen Diskurs I* (= Beiträge zur Fremdsprachenvermittlung, Sonderheft 24). Landau: Verlag Empirische Pädagogik. 2017. S. 217–240.

Rejzek, Jiří: *Český etymologický slovník*. Voznice: Leda. 2001.

Rocchi, Luciano: *Hungarian Loanwords in the Slovak Language III*. Trieste: EUT Edizioni Università di Trieste. 2010.

Tölgyesi, Tamás: *Lexikální germanismy v dnešní češtině. Studie kontaktovělingvistická*. Piliscsaba: PPKE BTK, zugegriffen am 01.03.2019, auf http://mek.oszk.hu/08400/08488/. 2009a.

Tölgyesi, Tamás: Středoevropská kuchyně jako zrcadlo kontaktů interkulturních. *Bohemica Olomucensia* 3, *Philologica Juvenilia*, 2009b, S. 145–150.

Tölgyesi, Tamás: Austriazismen in der mitteleuropäischen Küche. In: Szczęk, Joanna/Kałasznik, Marcelina (Hrsg.): *Intra- und interlinguale Zugänge zum kulinarischen Diskurs I* (= Beiträge zur Fremdsprachenvermittlung, Sonderheft 24). Landau: Verlag Empirische Pädagogik. 2017a. S. 271–281.

Tölgyesi, Tamás: Zur Herkunft des tschechischen Wortes *farkaš*. In: Janyšková, Ilona/Karlíková, Helena/Boček, Vít (Hrsg.): *Etymological Research into Czech. Proceedings of the Etymological Symposium Brno 2017,*

12–14 September 2017, Brno. Praha: Nakladatelství Lidové noviny. 2017b. S. 387–390.

Tölgyesi, Tamás: K původu slov farkaš a čabraka. Bohemica Olomucensia 10 – Linguistica, 2018, S. 196–203.

Tóth, Sándor János: The dominant language of bilingual speakers in South Slovakia. In: Lőrincz, Csongor/Görbe, Tamás/Halász, Hajnalka (Hrsg.): Berliner Beiträge zur Hungarologie. Schriftenreihe des Fachgebiets für Ungarische Literatur und Kultur an der Humboldt-Universität zu Berlin 20. Berlin: Humboldt-Universität. 2019. S. 25–39.

Vašek, Antonín: K lingvistické karpatologii v západních Karpatech. In: Alabán, František (Hrsg.): *Recepcia hodnôt v literatúre a jazyku. Zborník príspevkov z medzinárodnej konferencie (4.10.2000)*. Banská Bystrica: Univerzita Mateja Bela. 2001. S. 128–141.

Vig, István: *Sprachkontakte im Alpen-Adria-Raum*. Szombathely: Institut für Romanistik der Hochschule Berzsenyi Dániel. 2007.

WDW = *Wörterbuch der deutschen Winzersprache*, zugegriffen am 01.03.2019, auf http://190807.webhosting17.1blu.de/site/wdw/onlinewb/index. php?id=11890&c=.

Zaicz, Gábor (Hrsg.): *Etimológiai szótár. Magyar szavak és toldalékok eredete*. Budapest: Tinta Könyvkiadó. 2006.

List of Figures

List of Tables

Thomas Stolz & Nataliya Levkovych
On different ways of belonging in Europe

Luka Szucsich
Burgenland-Croatian as a contact language

Anna Tetereva, Viktoria Naukhatskaia
**Variation im Spracherwerb von Verben bei
bilingualen Kindern (Russisch – Deutsch)**

Linguistik International

Herausgegeben von Heinrich Weber, Wilfried Kürschner, Reinhard Rapp, Kazimierz Sroka, Jürg Strässler und Lew Zybatow

Band 10 Peter Kosta / Joanna Blaszczak / Jens Frasek / Ljudmila Geist / Marzena Xygis (eds.): Investigations into Formal Slavic Linguistics. Contributions of the Fourth European Conference on Formal Description of Slavic Languages - FDSL IV. Held at Potsdam University, November 28-30, 2001, Part I and Part II. 2003.

Band 11 Lew N. Zybatow (Hrsg.): Europa der Sprachen: Sprachkompetenz - Mehrsprachigkeit - Translation. Akten des 35. Linguistischen Kolloquiums in Innsbruck 2000. Teil I: Sprache und Gesellschaft. 2003.

Band 12 Lew N. Zybatow (Hrsg.): Europa der Sprachen: Sprachkompetenz - Mehrsprachigkeit - Translation. Akten des 35. Linguistischen Kolloquiums in Innsbruck 2000. Teil II: Sprache und Kognition. 2003.

Band 13 Stojan Bračič / Darko Čuden / Saša Podgoršek / Vladimir Pogačnik (Hrsg./ eds.): Linguistische Studien im Europäischen Jahr der Sprachen / Linguistic Studies in the European Year of Languages. Akten des 36. Linguistischen Kolloquiums in Ljubljana 2001. Proceedings of the 36th Linguistic Colloquium, Ljubljana 2001. 2004.

Band 14 Rolf Herwig (Hrsg./ed.): Sprache und die modernen Medien. Language and the Modern Media. Akten des 37. Linguistischen Kolloquiums in Jena 2002. Proceedings of the 37th Linguistic Colloquium, Jena 2002. 2004.

Band 15 Paweł Karnowski / Imre Szigeti (Hrsg.): Sprache und Sprachverarbeitung / Language and Language-processing. Akten des 38. Linguistischen Kolloquiums in Piliscsaba 2003 / Proceedings of the 38th Linguistics Colloquium, Piliscsaba 2003. 2006.

Band 16 Maurice Vliegen (Hrsg.): Variation in Sprachtheorie und Spracherwerb. Akten des 39. Linguistischen Kolloquiums in Amsterdam 2004. 2006.

Band 17 Tania Avgustinova: Language Family Oriented Perspective in Multilingual Grammar Design. 2007.

Band 18 Sabine Kupper: Anglizismen in deutschen Werbeanzeigen. Eine empirische Studie zur stilistischen und ökonomischen Motivation von Anglizismen. 2007.

Band 19 Franc Marušič / Rok Žaucer (eds.): Studies in Formal Slavic Linguistics. Contributions from Formal Description of Slavic Languages 6.5. Held at the University of Nova Gorica, December 1-3, 2006. 2008.

Band   20   Gerhild Zybatow / Luka Szucsich / Uwe Junghanns / Roland Meyer (eds.): Formal Description of Slavic Languages: The Fifth Conference, Leipzig 2003. 2008.

Band   21   Gerhild Zybatow / Uwe Junghanns / Denisa Lenertová / Petr Biskup (eds.): Studies in Formal Slavic Phonology, Morphology, Syntax, Semantics and Information Structure. Proceedings of FDSL 7, Leipzig 2007. 2009.

Band   22   Hans-Georg Müller: Adleraug und Luchsenohr. Deutsche Zwillingsformeln und ihr Gebrauch. 2009.

Band   23   Olga Souleimanova (Hrsg./ed.): Sprache und Kognition: Traditionelle und neue Ansätze. Language and Cognition: Traditional and new approaches. Akten des 40. Linguistischen Kolloquiums in Moskau 2005. Proceedings of the 40th Linguistics Colloquium, Moscow 2005. Herausgegeben unter Mitarbeit von Wilfried Kürschner, Nataliya Lyagushkina und Heinrich Weber. Edited in cooperation with Wilfried Kürschner, Nataliya Lyagushkina and Heinrich Weber. 2010.

Band   24   Beate Henn-Memmesheimer / Joachim Franz (Hgg.): Die Ordnung des Standard und die Differenzierung der Diskurse. Akten des 41. Linguistischen Kolloquiums in Mannheim 2006. Teil 1 und Teil 2. 2009.

Band   25   Gerhild Zybatow / Philip Dudchuk / Serge Minor / Ekaterina Pshehotskaya (eds.): Formal Studies in Slavic Linguistics. Proceedings of Formal Description of Slavic Languages 7.5. 2010.

Band   26   Markéta Ziková / Mojmír Dočekal (eds.): Slavic Languages in Formal Grammar. Proceedings of FDSL 8.5, Brno 2010. 2012.

Band   27   Maria Cristina Gatti: Spatiotemporality and cognitive-semiotic perspectives on corporate discourse for the web. 2013.

Band   28   Uwe Junghanns / Dorothee Fehrmann / Denisa Lenertová / Hagen Pitsch (eds.): Formal Description of Slavic Languages: The Ninth Conference. Proceedings of FDSL 9, Göttingen 2011. 2013.

Band   29   Evangelia Karagiannidou / Charis-Olga Papadopoulou / Eleni Skourtou (Hrsg./eds.): Sprachenvielfalt und Sprachenlernen: Neue Wege zur Literalität. Language Diversity and Language Learning: New Paths to Literacy. Akten des 42. Linguistischen Kolloquiums in Rhodos 2007. Proceedings of the 42nd Linguistics Colloquium in Rhodes 2007. 2013.

Band   30   Mareike Keller: Phraseme im bilingualen Diskurs. „All of a sudden geht mir ein Licht auf." 2014.

Band   31   Katrin Schöpe / Renate Belentschikow / Angelika Bergien / Armin Burkhardt / Karin Ebeling (Hrsg./eds.): Pragmantax II. Zum aktuellen Stand der Linguistik und ihrer Teildisziplinen. The Present State of Linguistics and its Sub-Disciplines. Akten des 43. Linguistischen Kolloquiums in Magdeburg 2008. Proceedings of the 43rd Linguistics Colloquium, Magdeburg 2008. 2014.

Band   32   Ewa Żebrowska / Mariola Jaworska / Dirk Steinhoff (Hrsg./eds.): Materialität und Medialität der sprachlichen Kommunikation. Materiality and Mediality of Linguistic Communication. Akten des 47. Linguistischen Kolloquiums in Olsztyn 2012. Proceedings of the 47th Linguistics Colloquium, Olsztyn 2012. 2014.

Band   33   Maria Grozeva-Minkova / Boris Naimushin (Hrsg.): Globalisierung, interkulturelle Kommunikation und Sprache. Akten des 44. Linguistischen Kolloquiums 2009 in Sofia. 2014.

Band   34   József Tóth (Hrsg./ed.): Die Sprache und ihre Wissenschaft zwischen Tradition und Innovation / Language and Its Study between Tradition and Innovation. Akten des 45. Linguistischen Kolloquiums in Veszprém 2010. Proceedings of the 45th Linguistics Colloquium, Veszprém 2010. 2015.

Band   35   Gerhild Zybatow / Petr Biskup / Marcel Guhl / Claudia Hurtig / Olav Mueller-Reichau / Maria Yastrebova (eds.): Slavic Grammar from a Formal Perspective. The 10th Anniversary FDSL Conference, Leipzig 2013. 2015.

Band   36   Ioana-Narcisa Crețu (Hrsg./ed.): *Quo vadis, Kommunikation?* Kommunikation – Sprache – Medien / *Quo vadis, Communication?* Communication – Language – Media. Akten des 46. Linguistischen Kolloquiums in Sibiu 2011. Proceedings of the 46th Linguistics Colloquium, Sibiu 2011. 2015.

Band   37   Markéta Ziková / Pavel Caha / Mojmír Dočekal (eds.): Slavic Languages in the Perspective of Formal Grammar. Proceedings of FDSL 10.5, Brno 2014. 2015.

Band   38   Juliane Besters-Dilger / Hana Gladkova (eds.): Second Language Acquisition in Complex Linguistic Environments. Russian Native Speakers Acquiring Standard and Non-Standard Varieties of German and Czech. 2016.

Band   39   Tanja Škerlavaj: Stilistische Funktionen von Mehrdeutigkeit in deutschen Werbetexten. 2017.

Band 40 Virginija Masiulionytė / Skaistė Volungevičienė (Hrsg.): Fremde und eigene Sprachen. Linguistische Perspektiven. Foreign and Own Languages. Linguistic Perspectives. Akten des 51. Linguistischen Kolloquiums in Vilnius 2016. Selected Papers of the 51st Linguistics Colloquium in Vilnius 2016. 2018.

Band 41 Anastasia Antipova: Die nationalsozialistische Sprachpolitik im besetzten Weißrussland 1941–1944. 2018.

Band 42 Lew N. Zybatow / Alena Petrova (Hrsg.): Sprache verstehen, verwenden, übersetzen. Akten des 50. Linguistischen Kolloquiums in Innsbruck 2015. 2018.

Band 43 Jürg Strässler (Hrsg./ed.): Sprache(n) für Europa. Mehrsprachigkeit als Chance / Language(s) for Europe. Multilingualism as a Chance. Auswahl an Beiträgen des 52. Linguistischen Kolloquiums in Erlangen (2017) / Selected Papers of the 52nd Linguistics Colloquium in Erlangen (2017). 2019.

Band 44 Luka Szucsich / Agnes Kim / Uliana Yazhinova (eds.): Areal Convergence in Eastern Central European Languages and Beyond. 2020.

www.peterlang.com