

In this chapter, I will outline the prospects of further research in this field. Examples provided throughout the present monograph are intended to exemplify the range of data to be included in the research of the lexical layers of cultural identity and the techniques of a consistent analysis of that data. They definitely do not build complete cultural-lexical profiles of any languages that were used and not all processes and features are characteristic for all Slavic languages. Neither do they represent an in-depth analysis of any discussed phenomena. I will therefore devote my attention here to research possibilities that can lead toward more definite accounts of languages and phenomena.

One obvious issue to discuss at the very outset are datasets and possible ways of improving them. I have repeatedly discussed research limitations stemming from the differences in datasets, most notably those that are brought about by different strategies and lexicographic solutions that authors of various Slavic dictionaries deploy. To overcome problems of this kind, it would be most useful to develop monolingual and bilingual lexicographic standards, which would lead to normalized data in comparing Slavic languages. It would be incumbent on Slavic studies centers and professional organizations, most notably the International Committee of Slavists (see Committee, 2017), to carry out such work. The benefits of such standards or guidelines would be multifold. They would enable cross-linguistic comparison of Slavic lexical data but, more than that, they would offer a platform for a consistent lexicographic treatment in all Slavic languages. Obviously, guidelines should account for the peculiarities of lexicographic traditions of each individual language. In addition to offering guidelines for lexicographic treatment, the standard should also encompass a standard for data representation (e.g., using the Text Encoding Initiative (TEI) scheme, version 5, see Burnard and Bauman, 2013, or Lexical Markup Framework (LFM); see Francopulo et al., 2007). Thus, for example, the guidelines could specify a consistent manner of stating in loanword dictionaries all languages in the chain of borrowing (e.g., the language of direct borrowing, further languages of origin). These guidelines should also specify appropriate descriptions for the language of direct, first indirect, second indirect, etc. etymology in the chosen data-representation standard.

The next possible contribution to creating more useful datasets could involve solutions that would normalize the existing anisomorphic data. The aforementioned guidelines could be useful to authors of future dictionaries. However, a wealth of data exists in various monolingual and bilingual Slavic dictionaries that needs to be normalized in order to make datasets for Slavic cross-linguistic research operational. The solutions in this field may include search-and-replace patterns that would appropriately tag the elements of the dictionary entry or extract lexical data from monographs. For example, if a dictionary states the language of direct borrowing using an angular bracket and the abbreviation for that language, e.g., [*lat.*], the search-and-replace pattern would look for this sequence and replace it with an appropriate tag, e.g. <etym type=borrowing><lang n=1>lat.</lang></etym>, etc.

I will now address the possibilities for further research that the proposed epistemological construct offers. Two general research directions offer particularly strong opportunities for the elucidation of lexical layers of cultural identity: lexical-cultural language profiles and in-depth contrastive studies. I will discuss them in turn.

Lexical-cultural language profiles offer the possibility of incorporating material from the rich body of literature of the dominant approaches to language and culture (most notably from linguistic culturology) into a broader and more systematic account of the lexical aspects of cultural identity. These profiles for each language and for their variants would specify general lexical ingredients that culturally determine the speakers of the given language or variant and the areas of variation between the speakers. Some of the features are overwhelmingly present in virtually all speakers, while others feature significant variation. To take an example of deep-layer carving of the conceptual sphere, in Slavic languages the lack of the distinction between *arm* and *hand*, *leg* and *foot*, *finger* and *toe*, etc. will almost universally be one of the profilers of cultural identity. On the other hand, the adjectives that describe different food items that have gone bad will be used with enormous variation – some speakers will use the equivalent of *bad* for all of them, while others will use specific designations such as *stale bread*, *sour milk*, *rancid butter*, etc. However, even in the latter case, the speakers will be culturally profiled by having a potential to use more specific terms. Potential is there even for those who have not mastered those adjectives, as they have a theoretical possibility to eventually master them. Similarly, the exchange layer may feature some sources of origin that generally define all speakers of the same language (e.g., Greco-Latin borrowings in all Slavic languages, which define the speakers as those from the European cultural circle), whereas other sources may exhibit strong geographical variation. Thus, Near Eastern loanwords not only profile Serbo-Croatian as similar to Bulgarian and Macedonian and distant from most other Slavic languages, but also mark a territorial identity of the speakers from the regions

that had the longest periods of Ottoman Turkish rule (especially Bosnia and Herzegovina, and Southern Serbia). These speakers are likely to know and use more of these words than those in other areas where the same standard language is used. In the surface layer, common to practically all speakers is the need to participate in negotiating standard language lexicon with elites (paying no attention to elites may even be a negotiation strategy). What varies from speaker to speaker are the kinds of attitudes they may have and what their level of compliance with the lexical norms is. In building lexical profiles, one could concentrate on more stable areas with less variation among the speakers, but then also list the areas with more variation which contribute somewhat less to their general cultural profile.

In-depth contrastive studies offer the possibility to focus on specific subject-matter fields in the deep and exchange layer and concrete norming maneuvers in the surface layer. Looking, for example, at fields such as common flora and fauna to examine the ratio of an inherited, a borrowed, and an engineered lexicon can reveal not only the differences stemming from the different environments of each Slavic language, but also differences stemming from cultural influences and geographical contacts. Similarly, if we look at the legal terminology in general use, we may be able to discover language-specific mental images and distinctions as well as those that are widespread in the Slavic realm and beyond it.

Another interesting field of research may be the relative availability of multiple equivalents. To exemplify this, I conducted a brief survey among the members of the Facebook group *Naš jezik* (devoted to Serbo-Croatian, with over 11,000 members in May of 2018, mostly from Belgrade, Serbia). I asked a question about the three Serbo-Croatian equivalents of the English word ‘uncle’: *ujak* ‘one’s mother’s brother’, *stric* ‘one’s father’s brother’, and *tetak* ‘one’s mother’s or father’s sister’s husband’. The question was formulated as follows. When I hear the word *дядя*, *Onkel*, *oncle*, *tío*, *tío*, *uncle*,¹ I think about: a. *ujak*, b. *stric*, c. *tetak*. Words before *uncle* were the same term in Russian, German, French, Portuguese, and Spanish, respectively. The survey was performed on February 18 and 19, 2018, and the respondents preponderantly chose ‘one’s mother’s brother’ (*ujak* 117 respondents – 91%, *stric* 9 respondents – 7%, *tetak* 2 respondents – 2%). Conducting research on the availability of lexical items from a broader and consistent dataset may reveal interesting facts. In this particular case, the three words for ‘uncle’ have the same frequency and they are similar phonologically and morphosyntactically. One possible explanation of this preponderance of responses selecting the word for ‘one’s mother’s brother’ may be a hypothesis about a higher prominence of mother (who in

¹ I.e. one German, French, Spanish, Portuguese, and English word for ‘uncle’ that covers all three Serbo-Croatian equivalents was provided.

a traditional family tends to home and children) – hence the choice of something that comes from the maternal side. To confirm it, one would need to select a whole lexical field and see if this factor is equally present in it.

I have already shown in Chapter 11 how the micro maneuver of using normative and cryptonormative labels can be studied in major Slavic monolingual dictionaries. Similar microanalyses can be done for macro maneuvers of enforcing the use of the standard lexicon and their micro maneuvers of stigmatizing lexical errors, giving normative advice. While all these may be present in all Slavic standard languages and their ethnic variants, the manner of their implementation may be very different, contributing thus to the cultural identity of each particular language in question. It seems that two grand ideological concepts play a pivotal role in these macro maneuvers: authority (using the tripartite model proposed by Weber, 1919) and nationalism (using the model initially proposed by Gellner, 1983). In particular, to use the three Weberian types of authority, while the authors of macro maneuvers always try to justify their authority as rational-legal, the facts on the ground may be different. It is hence interesting to see the extent to which these macro maneuvers rely on traditional authority (purist maneuvers seem to follow this pattern) and charismatic authority (especially in the various declarations of language councils, academies of sciences, and media appearances of prominent linguists). It is equally interesting to see, from a diachronic perspective, how much linguistic macro maneuvers have contributed to the formation of nations (using the aforementioned Gellnerian model of nation formation and the Brubakerian model of ethnicity).

An interesting possibility of diachronic research has been opened up by Vendina (2002) in her analysis of Old Church Slavonic as an expression of medieval mentality. Indeed, in each of the three layers the parameters change with the flow of time and an analysis of the lexicon of a distinct historical period may bring about insights into the cultural identity of that particular period in time. A related analysis may concern stability and change in time. Some of the parameters of cultural profiles are most resilient, others change abruptly. An interesting question is what contributes to resilience and what causes changes. Vendina (2014) announced another important possibility of looking into the distribution of lexemes in different areas of Slavic languages. Her analysis is based on dialectal data, but the same can be done with standard languages, concentrating on those lexemes that contribute to the cultural profiling of various speakers of Slavic languages.

One possible research direction lies in the field of corpus research. The examples throughout this book use lexical frequency in various dictionary datasets. This kind of lexical frequency can be called systemic frequency. Words and their forms also feature textual frequency, the frequency with which they are used in the texts of their language. To use a known example,

the lexical class of prepositions constitutes a rather modest proportion of the lexicon, but they are very prominent in the corpora of any language. For example, in Russian (according to Ljaševskaja and Šarov, 2009) there are 18 prepositions in the top 100 words and 4 of them are in the top 10 in this particular Russian corpus, which is certainly considerably more than their percentage in the lexicon. One possibility to expand the research proposed in the present monograph is to compare Slavic corpus data. Some of those data have already been included in the present research by using frequency data to come up with consistent lexical datasets in Chapter 4, but existing corpora of various languages enable the expansion of the current research. However, one should insert a word of caution here. Corpus datasets are considerably more problematic than their dictionary counterparts. Dictionaries might not perfectly represent the lexicon, but corpora are notorious in the variation of numerous criteria that can influence research. Take, for example, negative and positive characterizations of people, an example used in Chapter 5. The ratio of negative and positive characterizations will be drastically different if a newspaper corpus contains readers' comments. In that case the number of negative characterizations will generally be higher. If no readers' comments are present, then there is a lower number of negative characterizations, as the journalists are typically more guarded with such characterizations than the authors of the comments. This does not mean that corpus research on the lexical layers of cultural identity is impossible; it only requires sophisticated mechanisms for addressing the possible effects of non-representativeness. These mechanisms are certainly much more demanding than those deployed here for dictionary datasets.

In the concrete contexts of Slavic languages, most of them have relatively reliable corpora, and in some cases (e.g., with the Russian National Corpus, www.ruscorpora.ru) lexically and semantically tagged data and filtering tools enable quite sophisticated queries.

Using subjective frequency in addition to lexical frequency, addressed throughout the present monograph, and corpus frequency, a proposed area of expansion, may be a further area of meaningful data analysis. The speakers' subjective feeling of the frequency of some group of words (e.g., those stemming from a maneuver in the surface layer) may be not only indicative of their attitude but also decisive in their constant negotiation with linguistic elites.

I have already noted in Chapter 12 that attitudes, with their intensity, centrality, affective, cognitive, and behavioral components, represent an intricate cognitive construct and require fine-tuning of the research tools proposed here. This offers a distinct opportunity for interdisciplinary research with psychologists in conducting large-scale sophisticated surveys of all these attitude components and aspects. A further research prospect lies in the application of the models of attitude change to the study of lexical engineering and

refereeing maneuvers. In that light, the maneuvers would be seen as effective or ineffective mechanisms of attitude change.

To put it succinctly, the epistemological construct of the three lexical layers of cultural identity proposed in the present book offers the possibility to conduct in-depth analyses and to create general cultural profiles in a systematic and comprehensive manner, and thus to utilize the wealth of data that is available in various dictionaries and monographs.

Throughout the present monograph, I use examples from Slavic languages, which, as noted in the Foreword, offer ample material on lexical changes and various parameters of all three layers. However, this does not mean that the kind of analysis proposed here is restricted to those languages. In this sense, one distinct possibility lies in applying the research techniques demonstrated here (that are meant to be universally deployable) to the material of other languages. Obviously, some modifications will be needed in each group of languages and each contrasted language pair, but the general parameters should be useful in analyzing any standard language.

A further possibility may lie in the study of the cultures of urban and rural dialects or languages that have not been standardized. In such research expansion, the deep and the exchange layer would principally be used. There may also be cases of languages in the process of standardization where only some segments of the surface layer could be explored (for example, the cases of micro-languages where we only have the maneuvers of linguistic elites and the general body of speakers that remain unaware or uninterested in those maneuvers).

What I hope to have achieved here is the initiation of a new, more consistent, manner of exploring lexical layers of cultural identity. The proof of the pudding will be in potential research stemming from this first step and using the proposed methodology, modifying it where necessary, to produce an armamentarium for tackling the elusive links between language, culture, and identity.