

THE HISTORICAL DICTIONARY AS AN EXPLORATORY TOOL: A DIGITAL
EDITION OF VUK STEFANOVIĆ KARADŽIĆ'S *LEXICON SERBICO-
GERMANICO-LATINUM*

A Dissertation
Presented to
the School of Linguistic, Speech and Communication Sciences
Trinity College Dublin

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by
Toma Tasovac
2020

DECLARATION

THE HISTORICAL DICTIONARY AS AN EXPLORATORY TOOL: A DIGITAL
EDITION OF VUK STEFANOVIĆ KARADŽIĆ'S *LEXICON SERBICO-
GERMANICO-LATINUM*

Toma Tasovac

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

ABSTRACT

THE HISTORICAL DICTIONARY AS AN EXPLORATORY TOOL: A DIGITAL EDITION OF VUK STEFANOVIĆ KARADŽIĆ'S *LEXICON SERBICO- GERMANICO-LATINUM*

Toma Tasovac

My work on the digital edition of Vuk Stefanović Karadžić's *Српску рјечник: Lexicon Serbico-Germanico-Latinum* advances the theory and practice of digitizing historical dictionaries by demonstrating how detailed, annotated, electronic editions of legacy dictionaries can improve the scope of users' interaction with non-contemporary lexicographic data. I provide a substantial critique of the Function Theory of Lexicography as it pertains to historical dictionaries by spelling out how any theory of lexicography which reduces the dictionary — a complex socio-cultural *assemblage* — to a simplistic look-up tool ultimately limits its functionalities in the electronic medium. I propose a taxonomy of dictionary-specific information-related behaviors as a guiding principle for designing and implementing digital editions that fulfill the potential of the dictionary's *functional affordances*.

Starting with an exploration of TEI markup at the intersection of the text's linguistic properties (as an interpretative “model-of”) and its potential functionalities in the digital environment (as a prospective “model-for”), I develop the innovative concept of *dictionary slices* and *feature-based browsing* to expand the exploratory potential of remediated historical dictionaries. I articulate and justify the need for a medium-specific theory of electronic lexicography that will be firmly situated within the wider field of Digital Humanities, and sketch out some of the

main challenges facing the field in the years to come.

A TEI-encoded text of the *Lexicon-Serbico-Germanico-Latinum*, an application programming interface (API), and a published online edition are essential parts of this dissertation. Details on how to access them are provided in the appendices to the main text.

CONTENTS

ABBREVIATIONS	x
INTRODUCTION	1
1 WHAT IS IT AND WHAT DO YOU DO WITH IT?	16
1.1 How dictionaries came to be	17
1.2 Dictionary as a socio-technical assemblage	21
1.2.1 Dictionary as text	24
1.2.1.1 Typographic fixity: the legacy of print	24
1.2.1.2 Non-linearity: hypertext and écriture	28
1.2.2 Dictionary as a model of language	35
1.2.3 Dictionary as a tool	39
1.2.3.1 Dictionary as an ideological tool	42
1.2.3.2 Dictionary as a functional tool	45
1.2.3.3 Dictionary as a research tool	53
1.2.3.4 Dictionary as a digital tool	56
1.2.4 Dictionary as a cultural artifact	63
1.3 Affordances: the dictionary's actionable properties	69
1.3.1 The concept of affordance	70
1.3.2 Functional affordances	72
1.3.3 Types of information-related behavior	73

1.3.4	Doing things with and in dictionaries	78
2	AN OVERVIEW OF THE HISTORY OF ELECTRONIC LEXICOGRAPHY	84
2.1	DH and eLexicography: common roots	86
2.2	The sixties: a brave, cumbersome new world	88
2.3	The seventies: imagining the future	95
2.3.1	Technology as a double-edged sword	95
2.3.2	The missing infrastructure	102
2.4	The eighties: toward standardization of data formats	105
2.4.1	An emerging discipline	105
2.4.2	Common standards	109
2.5	Conclusion	112
3	DATA MODELING AS AN INTERPRETATIVE FRAMEWORK	113
3.1	Why model in the first place?	113
3.2	Functional encoding	118
4	<i>LEXICON SERBICO-GERMANICO-LATINUM</i>	128
4.1	A literary language for a 'bookless' people?	128
4.2	Typology of the dictionary	134
4.3	Importance of the dictionary	136
4.4	Different editions	138
4.4.1	First edition (1818)	138
4.4.2	Second edition (1852)	138
4.4.3	Incomplete third edition (1892)	139
4.4.4	Third edition (1898)	140
4.4.5	Fourth edition (1935)	141
4.4.6	Differences between editions	141
4.4.6.1	New content from E2 marginalia	141
4.4.6.2	Corrected lemmas	142

4.4.6.3	Deleted content	143
4.4.6.4	Accidental omissions	144
4.4.6.5	Marginalia from E1	144
4.4.6.6	E3-specific additions	144
4.4.6.7	Cross-references introduced in the third edition	145
4.4.7	Second Edition Reprint (1986)	145
4.4.8	Digital Edition (2017)	146
5	ENCODING <i>LSGL</i>	148
5.1	Encoding principles	148
5.2	Macrostructure	150
5.2.1	Entries	150
5.2.2	Homographs	153
5.3	Microstructure	155
5.3.1	General remarks	155
5.3.2	Forms	158
5.3.2.1	Lemma	159
5.3.2.2	Other word forms	159
5.3.2.3	Word endings	160
5.3.3	Grammar	161
5.3.4	Etymology	162
5.3.5	Usage	163
5.3.5.1	Types of marking	165
5.3.5.2	Diatopic	166
5.3.5.3	Diatextual	167
5.3.5.4	Diaphasic	170
5.3.5.5	Diastratic	171
5.3.5.6	Hint	172
5.3.5.7	Hint in German	173

5.3.6	Senses	173
5.3.6.1	Definition	175
5.3.6.2	Translations	175
5.3.6.3	Examples	179
5.3.6.4	Additional content	180
5.3.7	Names and named entities	181
5.3.7.1	Toponyms and anthroponyms as entries	181
5.3.7.2	Places	183
5.3.7.3	Persons	184
5.3.7.4	Vuk's self-reference	184
5.4	Mediostructure	186
5.4.1	Explicit cross-references	187
5.4.1.1	Cf. (confer)	189
5.4.1.2	Cf. added in E3	190
5.4.1.3	Cf. specific sense	191
5.4.1.4	Cf. specific sense added in E3	192
5.4.1.5	Cf. specific sense numbered in E3	194
5.4.1.6	Cf. with homographs	196
5.4.1.7	Cf. specific sense in a particular homograph	197
5.4.1.8	Cf. changed to vide in E3	197
5.4.1.9	Cf. with a foreign word: hidden etymologies	198
5.4.1.10	Vide	199
5.4.1.11	Vide specific sense	199
5.4.1.12	Vide particular homograph	200
5.4.1.13	Vide specific sense in a particular homograph	201
5.4.1.14	Vide changed to cf. in E3	201
5.4.1.15	Vide changed to dim. in E3	202
5.4.1.16	Variants	202

5.4.2	Implicit cross-references	204
5.4.2.1	Cursive references	204
5.4.2.2	Extended references	205
5.5	Editorial interventions	208
5.5.1	Marginalia from E1	208
5.5.2	Accidental omissions from E2	209
5.5.3	Marginalia from E2	210
5.5.3.1	Additions to existing entries	210
5.5.3.2	New entries	214
5.5.4	Vulgarisms from E1	215
6	IMPLEMENTING AN API-CENTRIC DIGITAL EDITION	217
6.1	Text as service	217
6.2	Raskovnik API: accessing lexical data	222
6.2.1	Searching	222
6.2.2	Browsing	224
6.2.3	Serendipitous discovery	226
6.3	GUI: affordances in action	227
6.3.1	Searching	228
6.3.2	Primary search	228
6.3.3	Secondary search	229
6.3.3.1	Cross-reference pop-ups	231
6.3.3.2	Entry collections	232
6.3.3.3	Cross-reference collections	234
6.3.4	Browsing	235
6.3.4.1	Alphabetic browsing	236
6.3.4.2	Feature-based browsing	240
6.3.4.3	Geobrowsing	245
6.3.5	Serendipitous discovery	248

CONCLUSION	251
APPENDIX I: RASKOVNIK API	258
APPENDIX II: FULL-TEXT OF TEI-ENCODED <i>LSGL</i>	264
BIBLIOGRAPHY	265

ABBREVIATIONS

DH	Digital humanities
E1	Караџић, Вук Стефановић (1818). <i>Српски рјечник, истолкован њемачким и латинскијем ријечима. Serbisch-Deutsch-Lateinisches Wörterbuch. Lexicon Serbico-Germanico-Latinum</i> . Беч: Јерменска штампарија. [First Edition]
E2	Караџић, Вук Стефановић (1852). <i>Српски рјечник истумачен њемачкијем и латинскијем ријечима. Lexicon Serbico-Germanico-Latinum</i> . Беч: Штампарија јерменскога намастира. [Second Edition]
E3-incomplete	Караџић, Вук Стефановић (1892). <i>Српски рјечник истумачен њемачкијем и латинскијем ријечима. Треће издање државно. Lexicon Serbico-Germanico-Latinum. Editio tertia publica</i> . I. Edited by Јован Бошковић. Београд: Штампарија Краљевине Србије. [Third Incomplete Edition]
E3	Караџић, Вук Стефановић (1898). <i>Српски рјечник истумачен њемачкијем и латинскијем ријечима. Треће (државно) издање, исправљено и умножено. Lexicon Serbico-Germanico-Latinum. Editio tertia (publica), emendata</i> . Београд: Штампарија Краљевине Србије. [Third Edition]
E4	Караџић, Вук Стефановић (1935). <i>Српски рјечник истумачен њемачкијем и латинскијем ријечима. Четврто државно издање.</i>

Lexicon Serbico-Germanico-Latinum. Београд: Штампарија Краљевине Србије. [Fourth Edition]

E5	Караџић, Вук Стефановић (2017). <i>Српски рјечник истумачен њемачкијем и латинскијем ријечима: дигитално издање. Lexicon Serbico-Germanico-Latinum</i> . Edited by Тома Тасовац. In Расковник: речничка платформа. Београд: Институт за српски језик САНУ и Центар за дигиталне хуманистичке науке.
FT	Function Theory of Lexicography
LSGL	<i>Lexicon Serbico-Germanico-Latinum</i> (in general)
LSP	Language for special purposes
TEI	Text Encoding Initiative

INTRODUCTION

A prehistory

I remember very vividly the first time I held Vuk Stefanović Karadžić's *Српску рјечник* [Serbian Dictionary] in my hands. I was around ten years old. My parents had moved several years before, but a portion of their books remained unpacked and forgotten in the basement. When my mother finally noticed that some books were missing, she asked me to help her get them. One of them was a rugged old copy of the 1935 edition of the *Lexicon Serbico-Germanico-Latinum*. It had my father's signature in the upper left corner of the first blank page, dated 1950, as a stamp of ownership, if not necessarily, as I would soon realize, actual use. It had a soft cover, which was distinctly yellow. And it smelled old. The spine of the book had definitely seen better days. Parts of it were missing, revealing both a layer of protective mashed fabric and horizontally aligned rows of stitches in the upper, middle and lower section of the spine. Because the book was Smyth-sewn, I could lay it completely flat on the desk, even though it was very thick. Lying open like that, the

book felt very smooth, like a fold-out map of sorts. The pages seemed remarkably clean. But some pages (I don't remember how many of them but definitely not just a handful) were still *uncut*. I thought this very strange. Why would anybody want to own a book like that? A book with hidden pages?

I didn't know at the time that books are printed on large, newspaper-sized sheets of paper, that these sheets are then folded several times, and sewn together, and that uneven trimming can leave some pages joined together both at the spine and at the edges. My mother was no expert at book binding either but she explained it to me that books "used to be sold that way." She handed me a paper knife and told me to cut away. Which I did. I sliced the pages open like there was no tomorrow. I was the first person *ever* to peek inside them. I was both exhilarated *and* terribly bored at the same time. And what I saw there didn't make much sense. I knew what a dictionary was. I knew that this was one. But it was not what I expected it to be. It was called *Српски рјечник*, i.e. *Serbian Dictionary*, but most of it was in foreign languages. I could read both Cyrillic and Latin at the time, but that didn't help much with the German, which was written in Gothic script. It felt like an ideal script for hiding one's secrets.

The Serbian parts of the dictionary were not written in the kind of Serbian that I knew. It was not the Jekavian dialect itself that was unfamiliar to me. The Jekavian aspect of it felt cosy and warm, close to home, or more specifically: close to my paternal grandmother, who had been born in Bosnia in 1903, and who kept her accent all her life, decades after moving to Belgrade in the thirties. She too would say *млијеко* /mlije:ko/ instead of *млеко* /mle:ko/. It was the wondrous, uncouth vocabulary itself, not the dialectal orthography, that seemed alien to me. Some words sounded right, but their meaning remained completely impenetrable. Others

sounded plain wrong: they were similar to the words I knew but seemed to have been somehow corrupted. Even when I recognized Serbian words that were printed in bold, they would often have other Serbian words appear next to them that seemed somehow off. Some felt ugly. Some felt positively funny. Most were strange.

But then there were stories. Stories of vengeful fairies and mighty dragons; of children getting devoured by witches, of magical plants that unlock all kinds of locks; of professional sluggards at Turkish court who had to prove they were fit for the job of doing nothing by refusing to stand up from a burning carpet; and stories of noisy village carnivals, cross-dressers and shape-shifters, of rainbows turning men into women and women into men. It was the strangest, the most mysterious and the most infuriating book I had ever seen.

I still own and use the very same copy.

Dictionaries and digital humanities

Dictionaries lie at the core of humanity's ability to conceptualize, systematize and convey meaning. Indeed, a dictionary is many things at once: a text, a tool, a model of language, and a cultural artifact deeply embedded in the historical moment of its production (Tasovac 2010). While it is true that we now live in the age of the electronic dictionary (de Schryver 2003), dictionaries have always played an important role in the interplay between production technology and knowledge taxonomies (McArthur 1986; Hüllen and Schulze 1988; Hüllen 1999). I believe this to be one of the main reasons why dictionaries hold such rich scholarly potential for the field of digital humanities (DH) (Schreibman *et al.* 2004; Moulin *et al.* 2011; Gold 2012; Berry 2012). Lexicography needs DH because DH — with its concern for the

abstract modeling of knowledge and its practical implementations in humanities research — can integrate and propel different trains of lexicographic and metalexicographic thought at the intersection of language and technology. But DH also needs historical lexicography because legacy dictionaries, with their highly dense structure and rich content, present a complex edge case for a field which is increasingly becoming focused on text-mining and visualizing unstructured, big data.

My work on a digital edition of Vuk Stefanović Karadžić's *Serbian Dictionary. Lexicon Serbico-Germanico-Latinum* advances the theory and practice of digitizing historical¹ dictionaries by showing how detailed, annotated, electronic editions of legacy dictionaries can improve the scope of users' interaction with non-contemporary lexicographic data. This thesis will attempt to answer three fundamental research questions. First, to what extent is the Modern Function Theory of Lexicography (FT) adequate as a vantage point for discussing historical dictionaries? Second, to what extent does FT account for the affordances of the historical dictionary in the digital medium? And, third, in what ways might digital editions of historical dictionaries develop mechanisms and interfaces which permit and encourage kinds of user interaction that go beyond those of simple look-up?

¹ The term *historical dictionary* is not unambiguous: it covers two different types of dictionaries and two different metalexicographic concerns (Haß 2011). On the one hand, it describes a type of dictionary that focuses on the historical development or a particular stage in the history of a language. On the other hand, the historical dictionary can simply refer to a non-contemporary lexicographic work. Vuk Stefanović Karadžić's 19th-century classic *Lexicon Serbico-Germanico-Latinum* falls into the latter category. In the rest of this chapter — and my dissertation as a whole — I will be concerned with this type of dictionary. I will use the term *historical dictionary* and the less common term *legacy dictionary* interchangeably.

In answering the first two questions, I will argue that any theory of lexicography which, like FT, reduces the dictionary to a simple look-up tool, without taking into consideration the complexities of the genre, a full range of its affordances, and the particular role that historical dictionaries play as research tools, ultimately limits the potential functionalities of the dictionary in the electronic medium. Inspired by and building upon these arguments, my answer to the third research question takes the form of an artifact that is simultaneously a demonstration, a case study, and, I hope, a culturally valuable output in itself, in line with the spirit and practice of the field of digital humanities: a set of encoding and design principles and, based on those principles, an implementation of a digital edition of a historical dictionary, which shows how structured data and semantic encoding can be successfully used for creating exploratory user interfaces of historical dictionaries. This cannot of course be an exhaustive answer to the question of the ways in which electronic lexicography can transcend look-up. Nonetheless, I hope that it will demonstrate some promising ways forward, and demonstrate conclusively that there are myriad ways in which the dictionary, in the electronic age, can be reconceptualized such that its value goes far beyond the merely functional.

Ultimately, this study as a whole points to the need for a new non-utilitarian, medium-specific humanistic theory of electronic lexicography. The theories of lexicography, as they developed in the second half of the twentieth century (and which I will discuss in Section 1.2.3 [pp. 39-63]) have to a large extent focused on the dictionary as a functional tool. Their goal was and still largely is to establish lexicography as a scholarly discipline with a rather limited scope and to distinguish it from other branches of knowledge, thereby increasing its academic or

cultural capital. This is not unusual of itself and can be viewed in the context of overspecialization that has been a prominent trend in the humanities and academia at large for quite a while (see, for instance, Epstein 2012). But this, I believe, has had a detrimental effect on both the intellectual scope of the discussions surrounding dictionaries and the way we implement digital editions of legacy dictionaries in the 21st century. The approach taken in this study is different perhaps because both my disciplinary point of departure and my angle of vision are different. As a humanities scholar working on legacy dictionaries in the digital domain, I would argue that a theory of lexicography cannot focus on dictionaries in purely instrumental terms and that it must embrace linguistic, historical, social, cultural and philosophical considerations as well as those coming from information science and other related fields.

This is also why this dissertation could not be anchored in a single disciplinary space, be it lexicography, linguistics or editorial scholarship. Digital humanities, as both an intellectual framework and a community of practice, embodies a particular kind of thinking about the human record in relationship to its technological manifestations. Digital humanities is very much about computational methods and building tools, but those methods and those tools are useful only if they help us provide new insights into our languages, cultures and histories. That's why my dissertation is not only a commentary on a particular digital edition of an individual historical dictionary, but rather a contribution to the way we think about what dictionaries *are* and what they can *do* in the digital domain. Regardless of the technology used to produce, disseminate and consume dictionaries, a truly *meaningful* future of dictionaries (as opposed to, say, automatic translation engines) will always be directly related to their essentially humanistic legacy.

My approach is based on the assumption that the field of lexicon retrodigitization should shift — just as it did in the broader field of digital libraries — from quantity and efficiency to quality, precision and user-friendliness; from breadth of electronic content and simple access tools to “sophisticated systems for ongoing use or apparatus providing interpretation” (Lynch 2002). The need to analyze, index and mark-up raw data and provide various types of annotations and metadata, named entities and other contextual information is of paramount importance for effective searching and retrieval of cultural heritage content (Borin *et al.* ; Borin *et al.* 2007; Schreiber *et al.* 2008; Christopher 2011). Historical dictionaries should be no exception.

While to the uninitiated, the question of digitizing dictionaries — or any other content for that matter — may appear as strictly technical and falling exclusively under the purview of software engineers and information science specialists, I will argue (in part against the views expressed, for instance, by Keating *et al.* 2010; Teehan and Keating 2010a; Teehan and Keating 2010b) that the functionality of a digital dictionary is contingent upon the purpose(s) we ascribe to it as editors or readers, and that text modeling is a complex scholarly activity that should *not* privilege the role of the software engineer above the other essential stakeholder — the humanist scholar. Ultimately, a digital edition is a “site for the exploration of knowledge”(Gabler 2006: 345). Which is why the priority of a digital humanist preparing a digital edition of a text should be, first of all, to analyze and understand the text; then model it and annotate it; and only then transform it into an interactive environment. Unless we approach the digitization of our legacy dictionaries in a painstakingly systematic and detailed manner — on the opposite spectrum from mass digitization efforts, plain-text data and simple lemma-based

indexes, we will hardly be able to advance beyond creating objects which are in terms of functionality and impact no more than digital surrogates of an already successful textual genre.

My dissertation shows how in a transdisciplinary field such as digital humanities, the very distinction between theory and practice is difficult to maintain as the more or less practical activities (such as text encoding, designing interfaces and creating interactive user environments), if done properly, require a significant intellectual investment that is both theoretically grounded and historically informed. Digitization is not a process of lossless conversion from one medium to the other: it is, above all, an opportunity for exploring, expanding and — if need be — reinventing our traditional scholarly objects.

The digitization boom that began in the mid-1990s dramatically altered the way we engage with immaterial cultural heritage. In this respect, lexicographic heritage has been no exception: it is by now well established that digitization can increase the use value of a historical dictionary, especially in global, networked environments (Kirkness 2008). Many projects have been initiated to create electronic editions of printed lexicons (see for instance Morrissey 1993; Lemberg *et al.* 1998; Christmann 2001; Fournier 2001). These retrodigitization efforts have been generally based on one of two approaches: either the production of “faithful” digital copies (at the cost of reproducing factual or typographic errors); or the structural modeling of the content, which treats the print edition as a data source rather than as an immutable text to be reproduced in its entirety (Lobenstein-Reichmann 2007). In either case, retrodigitization projects have usually not involved any significant amount of re-editing or expanding the actual dictionary content. What’s more, the interfaces of electronic editions of legacy dictionaries tend to be quite conservative.

In the field of retrodigitizing dictionaries — just as in the field of digital scholarly editions in general — the print page still rules supreme (see Earhart 2012).

This is not surprising. The goal of scholarly editing is to present a “reliable text” (MLA 2011) and there is a rich tradition of print-based scholarly editions that comes with distinct methodological and typographic conventions that have been proven over time (see Sahle 2013). The book form, with accompanying layout conventions such as line breaks, page breaks, various paratextual elements (prefaces, footnotes etc.) and bibliographic reference architectures, has become synonymous with humanistic scholarship. It would be mistaken, however, to treat the book interface as natural, transparent or non-technological. The conventions of the book may have shaped the transmission and preservation of textual and lexicographic heritage for centuries, which I will discuss at more length in Chapter 1, and especially Section 1.2.1.1 *Typographic fixity: the legacy of print* (pp. 24-28), but the fact of the matter remains that a page is as much an interface as a computer screen: it frames and mediates content.

MLA’s white paper *Considering the Scholarly Edition in the Digital Age* (MLA 2015) mentions interfaces as “environments for manipulation and exploration of the edition’s textual space and also as environments within which the user can occupy the role of a contingent editor, examining less-traveled editorial paths and their interpretive consequences” but stays away from explicitly addressing the more difficult question, which is: to what extent is the development of innovative exploratory interfaces in digital editions hampered by the legacy of print? While I fully concur with the formulation that the digital is not necessarily “a site of innovation” (Ibid.), I also can’t help but notice a trend of cautiousness in the editorial community that borders on the lethargic. Some of this caution is based on

the idea that a transcript, while convenient and easily searchable, is by definition inferior to a facsimile (Shillingsburg 2009), but this seems to me a moot point as it applies to transcripts in both analogue and digital forms. Nobody claims that we should use a transcript to study the materiality of a medieval manuscript, for instance. Gailey and Jewel, on the other hand, have expressed a concern about a ‘hipster ethos’ in the digital humanities community and “a trend that sometimes seems driven by innovation for innovation’s sake” (2012: 5). I do not understand what “innovation for innovation’s sake” could possibly mean or why we should not view technological innovation “as a value in itself.” If something is recognized as innovative by a community of scholars, then there has to be *some* inherent value in it. The motivation for such innovation seems to me to be of secondary importance. What matters *are* the results. Not “innovation for innovation’s sake,” but innovation for the sake of the humanities.

Regardless of what we see as the driving force behind digital humanities, we should always emphasize the fact that digitization is not a passive process of transposing content from one medium to another. It is much more a way of re-editing and re-interpreting the existing content. While this may be a generally accepted idea in many subfields of digital humanities, it has not been enacted in practice when it comes to digitizing legacy dictionaries.

Legacy dictionaries are often based on complex and not always entirely consistent structures. For this reason, they pose a particular technological challenge for text encoding, semantic markup, and database modeling (Fomin and Toner 2006; Nyhan 2006; Nyhan 2008; Mooijaart and van der Wal 2009; Lemnitzer *et al.* 2010). My dissertation shows, however, that the digital editing of legacy dictionaries, regardless of the technology used, is a complex lexicographic process that should

focus on the notion of the legacy dictionary as an exploratory tool. I see the process of digitizing historical dictionaries as the opposite of archiving, as an act of unforgetting.

Current retrodigitized historical dictionaries often fail to take a key reality into account: users interacting with a historical lexicon do not always have active command of the text's primary language. Even when a historical dictionary is retrodigitized with the user's needs in mind, the focus tends to be limited to easy-to-handle navigation and retrieval of elements from a full-text search (Christmann 2003); or on uniformization of existing data elements, such as dates (Kinable 2006). While these efforts are worthwhile and necessary — as they contribute greatly to editions that are more usable and efficient than their hardcopy counterparts — they don't change the fact that actual implementations of electronic legacy dictionaries remain, in essence, lookup tools (for words encountered in a given text) rather than exploratory tools (for unknown words or related concepts). This, I would claim, weakens both the scholarly and popular potential of the digitized historical dictionary. Precisely because historical dictionaries are research objects rather than consultation tools, their digital versions should be equipped with mechanisms for exploration and not only direct access.

To increase the scholarly value of our digital objects, we have to go back and go forward at the same time: go back to the considerations of lexicographic structures and functions which are, ultimately, I claim, indissoluble from the questions of semantics; at the same time, we need to go forward, explore, theorize and build dictionary interfaces which increase what Lobenstein-Reichmann (2007: 174) calls the “elektronische Mehrwert” [the electronic surplus value] of digitized historical dictionaries.

Outline

Unlike many European languages that underwent relatively uninterrupted linguistic and cultural evolutions, the standard Serbian language of today is partially a product of a caesura introduced by the publication of the *Lexicon Serbico-Germanico-Latinum*. By choosing to base his dictionary on the vernacular rather than the literary tradition of the previous epochs (Ивић 1990), Vuk Stefanović Karadžić created an unapologetically programmatic dictionary: a dictionary whose significance lies not only in the vocabulary it described, but also in the vocabulary it suppressed.

My dissertation project consists of a theoretical explication, a TEI-encoded text of the *Lexicon-Serbico-Germanico-Latinum*, an application programming interface (API) for the remote querying of the dictionary, and a published online edition.² Access to the TEI version of the dictionary and the API is explained in the Appendices. In the theoretical portion of my research, my main concern has been to establish how we can and why we should create better, more informative and more versatile editions of historical dictionaries for contemporary users. By using a hybrid text encoding approach that respects the integrity of the original text, but also takes advantage of the flexibility of the digital medium to provide supplementary annotations, I will explore, analyze, systematize and evaluate

² The digital edition of *LSGL* has been published on *Raskovnik: The Dictionary Platform* of the Institute of Serbian Language of the Serbian Academy of Arts and Sciences, <http://raskovnik.org>. The Institute hired the Belgrade Center for Digital Humanities, of which I am the director, to create the platform. The development work for the platform has been funded by the Serbian Ministry of Culture. I have been solely in charge of designing the system architecture, the data model, the API and the frontend functionalities, all of which have been directly informed by my work on this dissertation.

editorial techniques for creating modern, deeply-encoded, metadata-rich editions that serve not only as look-up mechanisms, but also function as exploratory tools for knowledge discovery. In this respect, my dissertation will contribute to the body of theoretical literature on text encoding and digital editions, while being firmly anchored in the more traditional scholarly literature on lexicography, linguistics and cultural studies.

In Chapter 1 *What is it and what do you do with it?* (pp. 16-83), I argue that the complexity of the dictionary genre is due not only to its informational density, but also to the fact that it is a *socio-technical assemblage*: an ordered system that is capable of subverting itself. I critique the Function Theory of Lexicography for its one-dimensional treatment of the dictionary as a simplistic, monofunctional information-extraction tool and I justify my notion of the historical dictionary as an exploratory tool against the notions of lexicographic access paths in electronic dictionaries. Finally, I apply the concept of *affordances* to explore the dictionary and its actionable properties in order to arrive at a more balanced set of principles that should be respected when creating digital editions of historical dictionaries.

In Chapter 2 *An overview of the history of electronic lexicography* (pp. 84-112), I look back at some of the key moments at the intersection of modern technology and lexicography. I focus exclusively on the early history of the use of computers in the production and study of dictionaries. Unlike the more recent developments, since the nineties onwards, the early history of computational or electronic lexicography is severely understudied. Yet some of the topics that were discussed in the early days continue to resonate today: how should we best structure and formalize lexicographic data so that we can process them efficiently and reliably? Are we developing methods, systems and technologies that can represent the complexity of

our data, or are we trying to fit the data into existing technical frameworks? When it comes to computers, is speed of access more important than the consistency and reliability of the data?

In Chapter 3 *Data modeling as an interpretative framework* (pp. 113-128), I contextualize the modeling of lexical data within the larger process of preparing digital editions. I analyze text encoding as both an interpretative abstraction of the source text and a projection of a future digital edition. Unlike Teehan and Keating (2010b), however, I argue for the primacy of semantic over “functional” encoding. A digital edition, like any tool, has to make its *affordances* visible and understandable to the user, but those should be considered as an additional data structure on top of the semantic encoding of primary sources. The reasons for this are both practical and conceptual. A semantically encoded text can serve as a scholarly tool of its own: encoding is, after all, a way of getting to know a text. One can, for instance, analyze the encoded version of a dictionary inside an XML Editor using XPath expressions without ever creating a proper digital edition as a publicly available point of access with an online graphical interface. Even more importantly, however, a semantically encoded text — one which is focused on the text itself, its structure and its meaning — can be more easily repurposed in various scholarly contexts than one whose encoding is custom-designed in view of the particular functionalities of a single digital edition.

In Chapter 4 *Lexicon Serbico-Germanico-Latinum* (pp. 128-147), I describe the importance of this particular dictionary as a milestone in the genealogy of the modern Serbian language and analyze the challenges that this dictionary, which went through several editions, poses for the editor of a digital edition. In Chapter 5 *Encoding LSGL* (pp. 148-216), I discuss the nuts and bolts of encoding the dictionary

in TEI, whereas in Chapter 6 *Implementing an API-centric digital edition* (pp. 217-250), I discuss the ways in which a deeply encoded historical dictionary can be used as a basis for developing innovative exploratory interfaces. Starting with an exploration of TEI markup at the intersection of the text's linguistic properties (as an interpretative "model-of") and its potential functionalities in the digital environment (as a prospective "model-for"), I develop the innovative concept of *dictionary slices* and *feature-based browsing* to expand the exploratory potential of remediated historical dictionaries.

Finally, in the concluding chapter (pp. 251-256), I draw a sketch of what I believe will be the most important developments in the field in the near future, including the need for articulating a medium-specific theory of electronic lexicography and the development of a pan-European lexicographic infrastructure.

WHAT IS IT AND WHAT DO YOU DO WITH IT?

I have already pointed out (on p. 3) that a dictionary is many things: a text, a tool, a model of language and a cultural artifact. It is not the case that some dictionaries are texts, and other tools; that some function as models of language, and some accumulate and promulgate cultural capital. A dictionary — every dictionary — is all these things at once: *a* list of many words, i.e. a singularity and a multiplicity; a map for finding your way *and* for getting lost; a shape *and* a shaper of language; a symbol of clarity *and* pure ideology.

The main goal of this chapter is, as the title suggests, two-fold: (1) to arrive at an all-encompassing theoretical conceptualization of the dictionary as a thing, which is to say a unified way of describing and understanding what it is as an intricate textual, linguistic, instrumental and cultural phenomenon; and (2) to move from this abstract conceptualization, which I will refer to as a *socio-technical assemblage*, toward a more practical appreciation of the dictionary's *functional affordances* as a set of inherent properties which make it possible for users to perform goal-oriented actions with it.

Neither of the two sets of issues discussed in this chapter require, strictly speaking, a chronological approach: this chapter aims neither to reconstruct a history of dictionaries as a potentially teleological narrative nor to set up a boxing ring for pitting print-based dictionaries against their digital counterparts. On the contrary, this chapter — and this dissertation as a whole — argues that a multifaceted, conceptually rich and theoretically sound approach to producing digital editions of legacy dictionaries today requires, more than anything, further probing into the very *idea* of a dictionary and into the *mechanics* of its use. Reductive approaches which are based on isolating individual aspects of the dictionary (most often, the dictionary as a simple look-up tool) inevitably result in digital products which limit the full aspect of the dictionary's functional affordances.

At the same time, a brief outline of how this thing called dictionary came to be can shed light on some of the conceptual complexities which will be addressed in Sections 1.2 and 1.3. What follows is, therefore, a selective genealogy with a conceptual twist: not the story of who wrote which dictionary and when, but how a dictionary became a thing in its own right.

1.1 How dictionaries came to be

The idea of the dictionary presupposes a degree of opacity in language. If language was transparent and immutable — fully masterable by a single person, perfectly understandable to each and every speaker, and immune to change over time — dictionaries would have had a much harder time establishing themselves as both practical tools *and* as comprehensive inventories of language. It should therefore come as no surprise that the oldest dictionaries that we know of — those produced

in the Middle East, Ancient China and Ancient Greece — were created to help their users understand language which was either foreign or archaic. The possibility of a dictionary was based on two mutually complementary ideas: (1) that words which are assumed to be known to the user can be used to describe those that are not; and (2) that words to be explained need to be arranged somehow in order for the user to be able to locate them.

A Sumerian-Akkadian bilingual glossary, compiled in the second millennium BCE and inscribed in cuneiform on twenty-four clay tablets containing around 9,700 entries, is the oldest surviving dictionary known to humanity (Lynch 2016). This simple word list, written in parallel columns, was organized thematically in groups starting with words describing legal and administrative matters, followed by those describing the material world (wooden objects, pottery, animals, parts of the body etc.) The work is often referred to as *Urra=hubullu*, based on its first entry: *urra* is the Sumerian equivalent of the Akkadian *hubullu*, meaning “debt, interest”. As a practical tool which was meant to help Akkadian scribes learn Sumerian, the *Urra=hubullu* had no grand poetic or philosophical pretensions, yet, by the very nature of being a selective compilation of words, it painted a picture, however rudimentary and however subjective, of an entire universe, “a comprehensive survey of the animate and inanimate world, geography, and stars, as well as artificially produced objects, victuals, and many other things.” (Soden 1994: 151)

In both Ancient China and Ancient Greece, dictionaries were explicitly related to rich literary traditions. By the time Homeric epics were written down, transmitted and consumed as works of literature³, their language was already

³ On the oral origins of *The Iliad* and *The Odyssey*, see Parry’s (1971a) and Lord’s (2000) groundbreaking comparative work on Homer and the formulaic tradition in South Slavic epic poetry.

perceived as difficult: the word forms and constructions of “archaic, Aeolic, Ionic, artificial, and possibly even ‘Achaean’ origin” (Parry 1971b: 7) merited explanatory notes or glosses (γλῶσσαι). As did the other works of literature, history or science written in particular Greek dialects (see Dickey 2007 and 2010). The Ancient Greek glossographic tradition culminated in the works of Aristophanes of Byzantium (c. 257- c. 180BC), the librarian at Alexandria, who wrote a series of titles such as *On words suspected of not having been said by the ancients*, *Of the names of ages*, and *On kinship terms*, although it is disputed whether these were self-standing works or parts of a comprehensive dictionary called *Λέξεις* (Slater 1976: 237, n 11); Apollonius Sophista and his synthesizing *Homeric Lexicon* in the first century CE, (Haslam 1994a; Haslam 1994b); and Hesychius of Alexandria, who, in the fifth or sixth century CE, compiled a lexicon of obscure poetic and dialectal words, phrases and short proverbs. Glossographic tradition was explicitly intertextual: it was very common for ancient lexicographers to cite previous authors and build upon each other’s work.

While the origins of Chinese lexicography can be traced back to pedagogic works such as *Historian Zhou’s Primer*, a widely circulated fifteen-volume textbook written by Zhou , the official historian of Emperor Xuan (827 bc–782 bc), for the purpose of teaching Chinese characters, the first comprehensive Chinese dictionary, aimed at explaining the meaning of words and expressions from early Chinese literature as well as providing encyclopedic information on a range of subjects from social and natural sciences, was the-so called *The Ready Guide*, produced by an unknown author around the third century BCE (Yong and Peng 2008). The widely accepted interpretation of the work’s title — “approaching what is proper, correct, refined”(Coblin 1993) — is a clear pointer to its didactic, prescriptive function.

Beyond the obvious formal challenges posed by the Chinese writing system, the most striking difference between Ancient Chinese and Ancient Greek lexicography was cultural: early Chinese lexicographic works were compiled on behalf of the government and used as textbooks in civil service examinations.⁴ As such, they were perceived as authoritative and official, an integral component of the state bureaucracy. It was only with the establishment of national academies from the sixteenth century onwards that European dictionaries would become part of a state-supported standardization infrastructure⁵.

Latin lexicography very much followed the Greek model: the first and greatest lexicon of Latin Antiquity, Verrius Flaccus' *De verborum significatu*, explicitly cited a number of early Roman authors, including some for whom it remained the only record of their work (Glinister *et al.* 2007). Unlike the Greeks, who do not seem to have been particularly interested in learning foreign (i.e. “barbarian”) languages, the Romans produced a number practical Latin-Greek and Greek-Latin glossaries meant for helping users in everyday situations (Kramer 2011). In medieval times, the bilingual glossaries increased in complexity and gained important scholarly functions: explaining the grammatical differences between the languages (*idiomata generum*, *idiomata generum verbi* and *idiomata casuum*) and providing literal

⁴ “In the early years of the West Han Dynasty, ‘the students are not allowed to take an official examination until seventeen. Only those capable of reciting more than nine thousand characters from [Historian Zhou’s primer] get the opportunity to be selected as candidates for government positions’” (Yong and Peng 2008: 47)

⁵ The Italian Accademia della Crusca was founded in Florence in 1583. The first edition of its *Vocabolario degli Accademici della Crusca*, which was published in 1612, established the notion of the academic dictionary as the arbiter of a national language. A similar stabilizing and standardizing approach towards the vernacular was later taken up by the French Academy and the Royal Spanish Academy.

translations of words or whole sentences from one language to the other (so-called *hermeneuma*) (Kramer 1996: 33-5).

Even though the above-mentioned early lexicographic works were produced under different historical circumstances and in very different cultural settings, they all seem to have grown out of a perceived need to overcome the difficulties posed by linguistic diversity, both in space and time. From their earliest days, dictionaries served as tools for facilitating communication and understanding: *synchronously*, between different dialects and languages; and, *diachronically*, by making historical changes in language easier to grasp. While the initial impulses to create glossaries and other types of lexicographic content were based on the so-called “hard-word” and etymological principles, lexicographers will, over time, adopt a more wholistic, documentary, approach to describing language. What was, originally, more or less, a collection of linguistic curiosities, will, with the spread of literacy and, eventually, print technology, morph into a different, maximalist genre, aiming to represent and contain *all* of language. In theory, at least.

1.2 Dictionary as a socio-technical assemblage

Most lexicographers and linguists will be happy to point out that a very common English phrase “to look something up in the dictionary” borders on grammatical and philosophical nonsense. Strictly speaking, there is no such *thing* as *the* dictionary. There are many dictionaries around but not a single one of them is the ultimate instantiation of the genre or the ultimate source of authority. And while it is important to recognize lexicographic authority for what it is — a claim to authority, rather than authority itself — it is equally important to identify

conceptually what the dictionary is. But to do that we need to think beyond lexicography as the “art and craft” of making dictionaries (Landau 1984).

The dictionary is a socio-technical *assemblage*. I use *assemblage* in the sense(s) ascribed to it by Deleuze and Guattari (1987).⁶ The dictionary (not as a Platonic ideal, but a principle, a set of relations, an epistemic possibility) is an *assemblage* in two very different senses: on the one hand, because as an ordered system (and, perhaps, not just any ordered system but *the* ordered system, a perfect territory), it very concretely manifests and represents the stratification of reason as such. On the other hand, it is a product of the *nomadic* spontaneity of language, “bearing a foreign tongue within each language as a whole” (514). It is not one or the other, or sometimes one, and sometimes the other. It is always both. Which is why a dictionary can be and has been for thousands of years a perfectly useful tool,

⁶ Deleuze and Guattari do not use the word *assemblage* in French. Rather, they use *agencement*. Nail (2017) has argued that *assemblage* is not a good translation because *agencement* is derived from the verb *agencer*, “to arrange, to lay out, to piece together” (*Le Robert Collins*), whereas the English *assemblage* means “the joining or union of two things” or “a bringing or coming together” (OED). Nail argues that a “layout or arrangement is not the same thing as a unity or a simple coming together” (22), and that this problematic translation can confuse English readers. This is an extremely tenuous argument, which would not be worth mentioning here, except that: a) it accidentally or purposefully distorts the meaning of one of the dictionary definitions; and b) it conflates the cultural authority of the dictionary with its practical use as a tool. When Nail says that a “layout or arrangement is not the same thing as a unity or a *simple* coming together” he is introducing a qualifier (“simple”) which is actually missing from the definition he is quoting. There is no “simple” in “a bringing or coming together” and there is no reason to read a far-reaching difference into “to piece together” and “to bring or come together” other than to prove a point one has already made. There is nothing wrong with an act of *interpreting* dictionary definitions, but one should do it carefully and with an understanding that definitions are interpretations themselves. Had Nail looked up *assemblage* in a different dictionary, for instance the *New Oxford American Dictionary*, he would have seen *assemblage* defined as “a collection or gathering of things or people” with no mention of “union”. After all, there are only so many different way in which one word can be defined. The dictionary is a powerful symbol, while a dictionary, no matter how good, and no matter how influential — only one, particular, imperfect instantiation of that symbol.

while also being a contradiction in terms: a text which is impossible to summarize, or retell, or truly understand.

The dictionary, like every assemblage, is “basically territorial” (Deleuze and Guattari 1987: 503). In principle, it delimits a territory (a language, a national identity, a state), but in reality, it shows that the limits are extremely porous: that words come and go, in and out of usage (fashion or style), from and to other territories (domains, dialects, languages). The dictionary is relational: it can be arranged and rearranged in an infinite number of ways without ever becoming one. Its only unity is “that of co-functioning: it is a symbiosis, a ‘sympathy’” (Deleuze and Parnet 1987: 69). If anything, although that’s no small matter, as we shall see later, it is or can become a “fragmentary whole” (Deleuze and Guattari 1994: 16). The dictionary, like an assemblage, is:

constituted by lines of deterritorialization that cut across it and carry it away...The territory is just as inseparable from deterritorialization as the code from decoding. (Deleuze and Guattari 1987: 504)

Because it is an assemblage, the dictionary can symbolize order and subvert that order at the same time. It can celebrate order *and* the utter randomness of that order.

In the rest of this chapter, I will look at different conceptual and disciplinary ramifications of the dictionary as a text (pp. 24-35), a tool (pp. 39-63), a model of language (pp. 35-39) and a product of culture (pp. 63-69). Each of these aspects — or strata — of the *assemblage* gives valid but necessarily partial answers to the question what the dictionary *is*. In the second part of this chapter (pp. 69-83), I will turn to the question of what one can *do* with a dictionary by exploring the notion of *functional affordances*, i.e. the dictionary’s actionable properties, in relation to

information-related behaviors (pace Bates 2002). The main goal of Chapter 1 is to demonstrate how the complexities of the dictionary genre and the possibilities afforded by the dictionary as an information-bearing object — possibilities which extend beyond targeted information extraction, i.e. looking up a concrete word — should be taken into account when discussing the remediation of legacy dictionaries in the digital age.

1.2.1 *Dictionary as text*

1.2.1.1 *Typographic fixity: the legacy of print*

Most dictionaries define the basic meaning of the word *dictionary* in a similar fashion: as a “book” (OED, Chambers) or a “reference work” (Duden, Wiktionary) containing a “list of words” and their explanations. On its surface, *dictionary* is a relatively simple lexeme to describe and most dictionaries do indeed define it either in terms of a physical object (a book) or more abstractly as a source of information (a reference work). The choice of *book* as the *genus proximum*, however, raises questions about how dictionaries are perceived and evaluated, not only by lexicographers themselves, but also by the general public. We live in an age in which the production, distribution and consumption of dictionaries is increasingly taking place online; in which dictionary portals are making it possible to search simultaneously across various lexicographic resources; and in which some publishers are altogether giving up on the print editions of their lexicons and encyclopedias. Yet the notion of the dictionary as a print product still seems to be firmly embedded in our lexicographic consciousness. Why is that?

Ascribing extra-linguistic value to individual dictionary definitions and drawing far-reaching cultural and philosophical conclusions from them is not a

procedure that should be taken lightly and without reservation. After all, dictionary writing is writing under duress or, as Fillmore (1988; quoted in Kilgarriff 2008: 148) put it, a lexicographer is a “lexicologist with a deadline.” The dictionary is a specialized genre constrained by the strict limitations of canonical, repetitive structures and a highly condensed mode of expression. Dictionary definitions are products of exigency, not poetic sentiment, let alone philosophical ambition. The specific constraints of the genre, however, do not obviate the more fundamental fact that a dictionary is, above all, a kind of text:

Le dictionnaire n'est pas seulement un objet, un produit de consommation, défini par des besoins socio-culturels, c'est aussi et surtout un *texte*, un *discours* continu et clos. (Dubois 1970: 35)

It consists of basic textual units — dictionary entries — and a complex system of textual pointers indicating how these fragments are mutually connected. It is fragmentary yet its entries nonetheless build a whole (“ein Textganzes”).⁷ Whether we perceive text in constructivist terms as a type of connected written discourse; as a social object with a history and rituals of use; or as a meaning potential to be realized in the interaction with the user, dictionaries start as being *written* and end as being *read*:

With the dictionary as text, one does not locate the semantic answer; one reads signifiers explicating other signifiers. Reading is a contingent

⁷ “Die Information jeder Artikelposition ist einerseits von derjenigen der anderen Positionen getrennt und hat insofern eine gewisse Eigenständigkeit; andererseits aber - und das ist wichtiger - bilden alle Positionen zusammen ein Textganzes; jede Einzelinformation wird deshalb nur aus dem Ganzen verständlich, sie wird überdies durch die Information anderer Positionen aspektuell ergänzt, so wie auch die Gesamtinformation eines Artikels nicht als Summe, sondern als Integral von Einzelinformationen zu verstehen ist” (Reichmann 1989: 63)

act...Reading involves *constructing meaning*, not absorbing it. (Frawley 1985: 11)

Even though most dictionary users tend to approach the dictionary in strictly functionalist and instrumentalist terms, as a source of objective, transparent, unmediated information, the textuality of the dictionary is not something that can be ignored. Where there is text, there is also the possibility of interpretation. That's why we can — and sometimes really should — explore, question and analyze dictionary definitions not only for what they say but also for what they *imply*. The OED's or Chambers' definition of the dictionary as a kind of book is precisely one such definition. Why *book* and why not *text*? Text is written discourse but it is not medium-specific: a text can be read from a clay tablet, a papyrus scroll, a parchment made out of prepared animal skin, a hand-written codex, a printed book or a computer screen. Choosing to say that the dictionary is a kind of book binds the notion of explication of lexical knowledge to one particular distribution medium with its own social and material boundaries. It prescribes the general look and feel of a dictionary, not to mention its physical and informational affordances.

Historically speaking, relating lexicographic works to the print medium makes sense because the growth and development of the dictionary as a textual genre was very much a product of the print culture as it started emerging in Europe from the late 15th century onwards. The “typographic fixity” made possible by the printing presses (Eisenstein 1979: I, 116-120) fully transformed the European intellectual landscape: its methods of data collection, retrieval and storage, and its patterns of communication. In many ways, it was the substantial, durable and reproducible form of the book — its perceived stability and palpable reality as

opposed to the ephemeral nature of speech — that made it possible for dictionaries to become authoritative and widely shared reference works in the first place:

Once these persons [grammarians] have codified their value judgments, the resulting grammars, dictionaries, and treatises are likely to be considered as having ontological priority, in part because of their material stability (Joseph 1987: 37)

Printing profoundly influenced the consolidation of national languages through the standardization of spelling, grammar and vocabulary (Febvre and Martin 1958/2010). The mechanical reproduction of “print-language” in turn, contributed to the creation of national consciousness by creating “unified fields of exchange... below Latin and above the spoken vernaculars” and helping “build that image of antiquity so central to the subjective idea of the nation” (Anderson 1991: 44). The encyclopedism of the 18th century and a proliferation of dictionaries and other forms of reference materials during the Age of Enlightenment was both an intellectual and a medial response to the growth of book publishing (Darnton 1979). After all, knowledge is always anchored in and shaped by the media of communication (McLuhan 1962; Innis 1951/1995). As the medial complexity of a culture rises, so does the need for classifying, organizing, interpreting and preserving the existing pool of knowledge (Hartmann 2003). In this respect, dictionaries and encyclopedias as synthesizing products of collective knowledge have offered a structured counterbalance to the overwhelming quantities of unstructured data — and they continue to do so to this even today, in the Age of the Internet (Lobenstein-Reichmann 2007).

As texts, dictionaries are secondary modeling systems: they outline the contours of language and make meaning classifiable, definable and — ultimately —

controllable. The notion of the dictionary as a book — *das Wörterbuch*, a wordbook — is not a terminological accident but rather a complex cultural concept that evolved over time. But the history of the *idea* of the dictionary — of what this peculiar textual genre is and what it represents — is far from over. That is why dictionary definitions of the word *dictionary* will have to remain provisional. This should not prevent us from asking broader questions about what the dictionary is — and, indeed — what it could become. What happens when a printed dictionary gets translated into an electronic dictionary? What is lost and what gained, if anything, in this process of translation? How does the transformation of dictionaries from discrete physical objects into networked hypertexts affect their cultural status? Do we need a new, medium-specific theory of electronic lexicography?

1.2.1.2 *Non-linearity: hypertext and écriture*

The dictionary is a text, but also a very particular kind of text, a prototypical *hypertext*: a network of nodes and links which, unlike traditional, sequential texts, are meant to be read in a non-linear fashion (see Delany and Landow 1991; Bolter 1991; Landow 1994; Cotkin 1996; Bolter and Grusin 1999). The hypertextual nature of dictionaries as texts is not dependent on the digital medium. A paper dictionary made up of thousands of mutually cross-referenced entries is as much a hypertext as an online dictionary in which HTML hyperlinks provide a way for the user to navigate the dictionary by clicking on them. Hypertext is a mode of structuring and organizing text, not a medium-specific implementation.

Because hypertexts imply an abandonment of “systems founded upon ideas of center, margin, hierarchy, and linearity” (Landow 1994: 1) and because the reader experiences hypertext “as an infinitely decenterable and recenterable system” (57), dictionaries as texts could also be linked to the deconstructivist notion of *writing* as

inevitable selective, circular, biased and contradictory. This has already been established by Frawley (1985):

words are defined in terms of other words, which are then defined in terms of other words, and so on. The circularity of the dictionary is a product of the dictionary as text. (10)

Deconstruction as a mode of reading rests on the postulates of Saussurean linguistics: every sign is a construct linking a signifier (a form, a sound) and a signified (concept, idea, image); neither the signifier nor the signified can mean anything on their own; all meaning is relational, created by differences (Saussure 1916/1981; Derrida 1997). For Frawley, a dictionary is like all texts: it is a product of *écriture* and a kind of writing whose “very existence as text calls itself into question” and which “deconstructs itself in active production and reception” (8). I would claim, however, that a dictionary is not simply like any other text. It is a paradigmatic text from the point of view of deconstructivist critique: a text *before* and *after* all other texts, a text containing and making possible all other texts. The dictionary deconstructs language and lays bare the logical aporias of meaning while appearing to be doing exactly the opposite: establishing, classifying and explicating words and their senses.

The possibility of constructing, deconstructing and reconstructing meaning out of an alphabetic dictionary has led to interesting literary experiments. For instance, members of *Ouvroir de littérature potentielle (Oulipo)* — a group of French writers and mathematicians interested in techniques of constrained writing — used dictionary-based word replacement algorithms as poetry generators. The rule S+7, originally proposed by Jean Lesucré in 1961, required that each noun in a poem be replaced by the seventh noun following it in a dictionary. “To see a World in a Grain

of Sand” would become, depending on the dictionary, something like “To see a Worm in a Grampus of Sandblast”(Alison 2006).⁸

It should also come as no surprise that one of the most celebrated postmodernist novels of the 20th century — Milorad Pavić’s *Хазарски речник: роман лексикон у 100,000 речи* [The Dictionary of the Khazars: A Novel Lexicon in 100,000 words, Павић 1984] — is written as a pseudo-dictionary. It is described as a reconstruction of the lost *Lexicon Corsi*, supposedly compiled by a Pole, Joannes Daubmannus in 1691, but destroyed in 1692. It is a collection of fragmentary, alphabetized, cross-referenced entries about dreams, love, memory and knowledge set against the background of the religious conversion of the Khazar people in the eighth century. The “original” dictionary is claimed to have described, using various sources and various points of view, the way a Khazar khan invited representatives of the three major religions, Christianity, Judaism and Islam, to interpret his dreams, promising that his people would convert to the religion that offered the most satisfactory interpretation of his dreams. In the novel, all three religions claim a “victory.” Immediately following the conversion, however, the Khazars are invaded

⁸ Compare this to “found poems” in corpus word-frequency lists. Even David (2005), in an essay which argues against post-modernist “deformative” criticism of Jerome McGann’s *Radiant Textuality* (2001) notices the accidental intelligibility and poetic quality of a sequence of fourteen words taken from the frequency list of a corpus of thirty British and American novels published around 1900:

Burning companion
(Heaven kissed ladies learn)
Leaves natural notice
O society, spread twice west! (79)

Dictionaries or lists in general, whether ordered alphabetically or by frequency, can generate meaning. We can call that meaning accidental, non-systemic, deformative, playful, algorithmic or however we want, but it is meaning nonetheless.

by a Russian prince, their empire destroyed, their people scattered. What starts as a dream, ends as a story of destruction.

Pavić's novel has been hailed as a triumph of postmodern literary techniques: as a true hypertext, a novel that repudiates Aristotelean linearity and makes readers active participants in the creation of multiple narratives (Cotkin 1996; Olsen 2000; Burkhart 2007). Its fragmentary nature has been analyzed, for instance, in terms of its implications for the questions of national identity and nationalism (Jordan 1994; Wachtel 1997; Aleksić 2009) or epistemology and desire (Hayles 1997; Leitner 1998; Longinovic 1998). The fact that Pavić's novel is structured like a dictionary is usually interpreted metonymically: as a stand-in for a more general aesthetic and ideological point of view. *The Dictionary of the Khazars* is, after all, a work of literature. In the context of my research, however, this novel is important because it delineates the very contours of lexicographic textuality as such, of the dictionary as discursive space and a contradictory source of authority on language and the world. Pavić's novel reveals how lexicographic order — or what I will later call *infinite finitude* (see p. 65) — can be used to deconstruct that very order.

The novel is published in two editions (male and female) which are almost identical except for one “crucial” paragraph; it is divided in several sections: preliminary notes, dictionaries proper (containing three independently alphabetized ‘books’: “The Red Book” (Christian sources on the Khazar Question), “The Green Book” (Islamic sources) and “The Yellow Book” (Jewish sources), as well as two appendices, a postscript and a list of entries.

In the Preliminary Notes, the narrator encourages different modes of reading the text: looking up individual entries, reading the dictionary as a whole

(“from the beginning to the end”), from right to left and left to right, by sections (the three dictionaries proper), diagonally („дијагонално”) in triplets („по тројкама”) i.e. by selecting first those entries that are covered in all three sources (and marked up with a small triangle), or serendipitously, starting with the Book which opens itself by chance when the dictionary is opened (lit. unfolded) („најпре она која се прва отвори приликом расклапања овог речника” [19]):

Али, корисника овог речника не треба да обесхрабре ова подробна упутства. Он може мирне душе да прескочи све ове уводне напомене и да чита као кад једе: читајући, десним оком може да се служи као виљушком, левим оком као ножем, а кости да баца за леђа. И то је довољно. Може му се, додуше, десити да залута и да се изгуби међу речима ове књиге, као што је Масуди, један од писаца овог речника, залутао у туђим сновима и није више никада нашао пут натраг. (19)

[The user of this dictionary, however, should not be discouraged by these detailed instructions. He can easily skip all these preliminary notes and read as he eats: reading, he could use his right eye like a fork, and his left eye, like a knife, throwing bones behind his back. And that would be enough. Admittedly, the reader could lose track and get lost among the words of this book, the same way that Massudi, one of the writers of this dictionary, lost his way in other people’s dreams and never found his way back.]

The “lexicon novel” does not have to be read in its entirety: „из ње се може узети пола или само део, и на томе се може остати, као што се с речницима најчешће и чини” [one can take only one half or a part of it, and leave it at that, *as one usually does with a dictionary*, emphasis mine].

Pavić’s novel is not like a dictionary. It is a dictionary. Which means that it can be read “in an infinite number of ways” (18). But not only that:

То је отворена књига и кад се склопи, може се дописивати: као што има свог негдашњег и садашњег лексикографа, може стећи у будућности нове списатеље, настављаче и дописиваче. Она има одреднице, конкордансе и натукнице као свете књиге или укрштене речи и сва имена или појмове обележене овде малим знаком крста, полумесеца, Давидове звезде или другим знацима, треба потражити у одговарајућем речнику овог речника да би се нашло подробније обавештење о њима. (Павић 1984: 18)

It's an open book and when it is closed, it can be added to: just as it has its own former and present lexicographer, so it can gain new writers, continuers and annotators. It has entries, concordances and indices like holy books and crossword puzzles, and all the names or concepts marked up with a small sign of cross, half-moon, David's star or other signs should be looked up in the corresponding dictionary of this dictionary in order to find a more exhaustive information about them.]

Throughout the novel's Preliminary Notes, Pavić highlights the “openness” of the dictionary — both in the sense of the fact that it can have multiple compilers and the sense that it can be read in multiple ways. On a more formal level, however, Pavić exemplifies in the above passage the paradoxical limits of that openness, which is practically impossible to reflect in the English translation. In the following paragraph, I will try to explain how Pavić's choice of polysemous words or synonyms undermines, even at this micro-level, the proclaimed exhaustiveness of the dictionary as a textual genre.

When Pavić describes the act of “closing” an “open” book, he uses the polysemous verb *склопити*: *склопити књигу* is a common collocate, which means to close a book, but *склопити* and its nominalization *склоп* are primarily used in Serbian to denote the process of folding, or more generally construction or

assemblage.⁹ For Pavić's narrator, the dictionary seems like an open book not only in the sense that multiple lexicographers can work on it, but also in the very sense of how it is put together so that it can be read in multiple ways. The polysemy of *склопити* is essential here especially because Pavić chose *not* to use the direct Serbian antonym pairing of *open* and *close* (отворити - затворити).

Speaking of the dictionary's former and present lexicographers, Pavić's narrator says that it can also acquire future *списатеље* (writers), *настављаче* (continuers) и *дописиваче* (lit. those who add in writing). *Настављачи* and *дописивачи* are for all intents and purposes synonyms: the former noun denotes those who continue something (< *наставити*, to continue), and the latter those who add to something in writing (< *дописати*; prefix *до* + *писати*, to write). Even more surprising is his choice of words when describing the contents of the dictionary with its *одреднице*, *конкордансе* и *натуркице* (entries, concordances and indices [lit. entries]): in Serbian, *одреднице* and *натуркице* are full-fledged synonyms, used interchangeably to describe dictionary articles.

Pavić's lexical choices in the above passage offer a subtle qualification of the ideas expressed in it: the dictionary may be an "open" book, which can be read in an "infinite number of ways", but it is also, in a different sense, a "closed" system: a finite collection of signifiers, some of which point to different signifieds (as in *склопити*), while some signifieds, irrationally perhaps, can be expressed through multiple signifiers (*одреднице*, *натуркице*). By using two synonymous lexemes in the very same sentence to describe the concept of dictionary entries, textual items essential to lexicographic enterprise, Pavić highlights the problematic notion of

⁹ In a previously quoted passage, Pavić's narrator uses *расклопити*, the antonym of *склопити*, to describe a reading technique based on the random opening (lit. unfolding) of the book.

lexicographic infinity. The infinity of a circle is not the same as that of a straight line.

Even though the reader in Pavić's novel may be choosing his or her own narrative paths through the text, the novel in the shape of a dictionary exists not in order to help users find the one, true and only meaning of the novel, but instead to give them a chance to get lost while trying. While this may, at first, seem to be no more than an expression of literary fancy, it will have consequences for the way we use or consult a dictionary as a tool (see Sections 1.1.3 *Dictionary as a tool* [pp. 39-63] and 1.2.3 *Affordances: the dictionary's actionable properties* [pp. 69-83]).

1.2.2 *Dictionary as a model of language*

The entire plot of a 2008 episode of the US sitcom *Sarah Silverman Program* entitled "Kangamangus" revolves around the desperate attempt of the show's self-absorbed protagonist to leave a mark in history by inventing a slang word (*ozay*) that would "find its place among the people" and become popular (Schrab 2008). This, Sarah explains to her friend Brian — the author of a rival slang word, *dot-nose* — would be her "only shining moment in a lifetime of unshining moments." Her hopes, however, are shattered when she witnesses the arrival of a Sir Corin Ashley, the editor-in-chief of the *Oxford English Dictionary*, who is "delighted to inform" Brian that he and his colleagues have "decided to include dot-nose as a word within the English language." The episode ends with a pretentious "Oxford English Dictionary Word Induction Ceremony" in which Brian receives an award and Sarah gives her ambiguous concession speech:

Tonight, dot-nose became a word in our dictionary. Well, let's not forget that once, so did *holocaust* and *diarrhea*. One of them happened. And one of them continues to happen. (Ibid.)

The comic twists of the plot in this episode exploit the popular perception of the dictionary as the ultimate judge of socio-linguistic acceptability: if a certain lexical item is found in a dictionary, it is considered to be a legitimate part of a given language. As I already mentioned in the previous section, the fixity of the print medium has greatly contributed to the establishment of lexicographic authority. But there is also something else at stake here, which doesn't have so much to do with the question of the medium, but is rather related to the dictionary's topographical organizational structure. Once language is represented a prearranged and finite list of items, "we cannot avoid thinking of the language as a set of rules; for ordinary speakers, the idea of 'rule' may be expressed as 'correctness'" (Dolezal 2009: 696).

Lexicographers like to make a distinction between descriptive and prescriptive dictionaries: the former generally describe language as it is actually spoken and written in a certain linguistic community; while the latter tend to pass judgements on how a certain language *should* be spoken and written according to some institutionalized standard (see Wells 1973; Béjoint 2000; Svensén 2009; Béjoint 2010). Even though a theoretically sound, clear-cut and non-ambiguous distinction between descriptive and prescriptive aspects of polyfunctional dictionaries may be difficult to maintain, cultural and socio-political circumstances have led to the widespread perception of all dictionaries as authoritative arbiters of what constitutes language. That is why the speakers of a written language with a lexicographic tradition often use dictionaries as a source of authority to "settle family arguments" (Summers 1988), and also why legal arguments sometimes involve, for better or worse, the use of language dictionaries (see Liptak 2011; Brudney and Baum 2013).

Even more fundamentally, the dictionary presupposes a specific model of language as a unified, delimited system consisting of discrete meaningful units (words) which have basic, fixed meanings (Seargeant 2011). When the supposed editor of the OED on the Sarah Silverman Program declares that a word has been “included as a word *within* the English language,” (emphasis mine) he not only asserts the role of the OED as an instrument of lexical legitimization, but also postulates that the English language is an entity with clearly delimited borders, a privileged inside and an excluded outside. In many cases, identifying that one word “belongs to” a certain language (*cat* is English in a way that *chien* is not) is relatively easy in comparison to answering the question about what English — or any other language — is, where it begins and where it ends. After all, a named language is not an empirical entity: it is an abstracted construct based on socio-linguistic patterns of use. Even though a dictionary of a named language is usually a lexical record of a particular language variety aimed at a particular audience, by virtue of metonymy, it is typically perceived as a symbolic representation of the imaginary “whole” language.

The very idea of the dictionary and its organizational structure (entries containing ordered, numbered senses) has been predicated upon the belief that meanings are fixed and well-defined regardless of the context in which they occur. There are studies — especially those based on corpus analysis¹⁰ — which seem to suggest that the way dictionaries tend to disambiguate senses, resolve ambiguities

¹⁰ Corpus linguistics has made a dramatic impact on lexicographic practice by shifting the focus toward the observation of linguistic patterns such as collocations and colligations (Sinclair 1991; Fellbaum 2009) as well as looking for computational evidence about issues of semantic prosody (Sinclair 1987; Louw 1993; Stubbs 1995), thus challenging the position of the word as a privileged unit of meaning.

and avoid redundancy is “at best superficial and at worst misleading” (Hanks 2008: 125) because they create a false picture of what happens when language is used.¹¹ According to Hanks, meanings are events, not entities, and words consists of probabilistic and prototypical semantic components which may or may not be activated by the speaker in any given instance of use:

The corpus contains traces of meaning events; the dictionary contains lists of meaning potentials. (Hanks 2008: 130)

Psycholinguistic experiments involving the study of semantic priming also seem to suggest that the resolution of lexical ambiguities is a dynamic process which is dependent on both the dominant meaning of a word and the clues provided by the context in which the word appears (Williams 1992; Klein and Murphy 2001; Klepousniotou *et al.* 2008). Dictionary entries, with their predisposition for static, clear-cut, context-free, non-overlapping senses described in lexicographic shorthand can hardly approximate that process.¹²

Regardless of whether the dictionary is based, like Johnson’s, on the authority of “writers of classical reputation or acknowledged authority”, or, as is the case in modern, corpus-based approaches, on the statistically representative usage

¹¹ For a useful overview of approaches to practical word sense disambiguation and theoretical lexicology, see Kilgariff 2008.

¹² Alphabetic order is not the only possible organizing principle for lexical data. In 1985, a group of linguists and psychologists at Princeton University began developing a reference aid that would help users search for information in traditional dictionaries conceptually rather than alphabetically (Miller 1985). As the work proceeded, the authors transformed their initial idea into a more ambitious project for a new dictionary that would overcome the limitations of sense-defying alphabetization (1993). The result was WordNet, a comprehensive, open-sourced, machine-readable lexical database of the English language inspired by psycholinguistic theories of the human ‘mental lexicon’ (Singleton 1999; Aitchison 2003; Bonin 2004). Today, WordNets exist for many different languages and are widely used in various NLP tasks.

patterns of broader language communities, the lexicographer continues to play an essentially mediating role. The lexicographer selects the source(s) of authority, extracts the information from those authorities and interprets the information for the user:

A lexicographical definition, we shall argue, does not in most cases identify a meaning independently existing in actual usage and discovered there by the lexicographer: it is deliberately constructed and allocated by the lexicographer on the basis of materials selected for study, and its allocation will depend on the viewpoint the lexicographer has chosen to adopt. (Harris and Hutton 2007: 78)¹³

The characters on the Sarah Silverman Program have a skewed picture of how words “get into the dictionary”. That picture, however, is part of the dictionary as the socio-cultural imaginary and is based on the fundamental ways in which dictionaries shape the common perception of what language is and how it functions. Dictionary editors, as a rule, do not make house calls to congratulate people on their neologistic prowess; there are no such things as word induction ceremonies; and, most importantly, the question of whether something is a word or not cannot always be settled once and for all by looking it up in a dictionary. Nonetheless, dictionaries remain powerful models of language, and lexicographers arguably its most influential modelers.

1.2.3 *Dictionary as a tool*

I already discussed (on p. 24) the conceptual difference between the two common ways of defining the dictionary — as a kind of book (i.e. a physical object); and as a

¹³ More broadly, Harris and Hutton view semantics as “the study and practice of human attempts to impose some degree of communicational determinacy on signs. The successes, failures and limitations of such efforts are, in our view, central to the enterprise of definition” (viii).

(more abstract) reference work. In Section 1.2.1.1, I argued that the continued conceptual association of the dictionary with the print medium can be attributed to the revolutionary and lasting role which the printing presses played in the development of modern scientific production, in general, and dictionary-making, in particular. In this section, I will focus on the notion of the dictionary as a *tool* and the way this instrumental aspect of the dictionary has been exploited in lexicographic theory and practice.

Simply put, a tool is an instrument used to perform a particular type of action.¹⁴ A tool is a tool if it can be used with the goal of effecting a particular result. A device which transfers physical force to modify or effect other objects or the surrounding environment (a knife, a hammer or a spade, for instance) is as much of a tool as an information retrieval mechanisms (a book index, a dictionary or a library catalogue) or a piece of computer software (a web-browser, a word processor, a calculator): their level of complexity may differ, and so can their situation of use, but they are all made to satisfy certain needs and allow their users to pursue certain goals. The essence of a tool lies in its readiness to be used.¹⁵

The idea of the dictionary grew out of a perceived need, not a theoretical concept (see Wiegand 1998; also my discussion of the perceived opacity in language

¹⁴ The OED supplements Samuel Johnson's literal definition of the word as '[a]ny instrument of manual operation' with a more explicit explanation: "a mechanical implement for working upon something, as by cutting, striking, rubbing, or other process, in any manual art or industry" (OED). Figuratively speaking, however, a tool has much broader connotations: "[a]nything used in the manner of a tool; a thing (concrete or abstract) with which some operation is performed; a means of effecting something; an instrument" (ibid.)

¹⁵ According to Heidegger, contemplating tools is the wrong way of going about understanding them: "the less we just stare at the hammer-Thing, and the more we seize hold of it and use it, the more primordial does our relationship to it become, and the more unveiledly is it encountered as that which it is—as equipment" (Heidegger 1962: 98)

as the initial catalyst for the “invention” of the dictionary in Section 1.1 [pp. 17-21]). Because dictionaries are tools, it should come as no surprise that they long predate lexicographic theory. This is not to say that lexicographic practice in its pre-theoretical phase was lacking a solid intellectual base. The “hard word” principle or the etymological principle, to name just two important pre-theoretical approaches to dictionary making (Hayashi 1978), were sufficient frameworks for creating reference tools with real, practical value: those that could be used for finding information about words, their meaning and their origin.¹⁶ But, as we already saw in Section 1.2.1.2 (pp. 35-39), the dictionary is more than a list of words: it is a model of language as such. It is, like all models, inherently selective: a limited and abstracted representation of a much larger, empirical entity. Because of this fact, a dictionary can be and has been used, very successfully, as a standardization tool: a source of ideological authority on what counts as good, proper or correct language. Any serious theoretical discussion of dictionaries which aims to be comprehensive needs to take into account the fact that the “toolness” of the dictionary is both functional and ideological.

The sections that follow will address four different ways in which we can speak of the dictionary as a tool: in Section 1.2.3.1 (pp. 42-45), I will discuss the notion of the dictionary as an *ideological tool* by analyzing the puristic concept of the dictionary as described in Samuel Johnson’s *Plan of a Dictionary of the English Language* (Johnson 2017). In Section 1.2.3.2 (pp. 45-53), I will address the notion of the

¹⁶ Hausmann’s *Kleine Weltgeschichte der Metalexikographie* points to five different types of sources of metalexigraphic thought before the First World War: prefaces to dictionaries, dictionary reviews, specialist articles in existing dictionaries and encyclopedias, relevant sections “in large scholarly manuals” [in großen wissenschaftlichen Handbüchern] as well as independent metalexigraphic works [eigenständige Metalexikographie] (1989b).

dictionary as a *functional tool* by analyzing the main tenants of the so-called Modern Function Theory of Lexicography. In Section 1.2.3.3 (pp. 53-55), I will focus on the notion of the dictionary as a *research tool* — a topic which has remained unexplored by proponents of the Modern Function Theory of Lexicography, but which is eminently important in the context of this dissertation and my work on digitizing legacy dictionaries. Finally, in Section 1.2.3.4 (pp. 56-63), I will look at the dictionary as a *digital tool* in order to discuss the implications of the medium in which users consult lexicographic material on the way we theorize the dictionary as such.

1.2.3.1 *Dictionary as an ideological tool*

While the English never created a national institution à la Accademia della Crusca or the Académie française — venerable institutions set up in order to oversee and regulate the development of the national language — a group of printers and booksellers in the 1740s decided to sponsor a new English dictionary that would perform some of the same functions as the famous academic dictionaries on the continent. They chose Samuel Johnson, a relatively well-respected but not particularly successful writer at the time. The exact reasoning for their choice remains unclear (Lynch 2009).

Johnson's *Plan of a Dictionary of the English Language* (2017) sketches out the contours of a highly ambitious lexicographic project whose "chief intent" is "to preserve the purity and ascertain the meaning of our English idiom" (Ibid.: 177). This is a striking turn of phrase for two reasons: (1) it positions the prescriptive quest for purity in language as if *before* the very act of semantic description; and (2) Johnson's use of the personal determiner *our* in "our English idiom" signals the value he ascribes to language not as an abstract but rather as a very concrete social entity, a community of speakers that share a common identity. Despite his self-

deprecating references to dictionary writing as “drudgery for the blind” and “artless industry” which consists of “beating the track of the alphabet” while requiring no personal quality other than “dull patience” and “sluggish resolution” (Ibid.), Johnson’s actual theoretical projection is based on the vision of lexicography as an art which is both meticulously methodical and monumentally moralistic. Almost everything he says he wants to accomplish in and with his dictionary as a tool — an object whose value “must be estimated by its use” (Ibid.) — is also framed by the singular mission of pursuing the ideal of linguistic purity as a way of boosting national pride and the “the reputation of our tongue” (Ibid.: 188).

Central to this type of lexicographic tool-building enterprise is the process of word selection. Even though “[i]t was not easy to determine by what rule of distinction the words of this dictionary were to be chosen,” (Ibid.: 177) Johnson’s job seems to have been made easier by a forcefully articulated ideological stance that can most succinctly be described as linguistic nativism. A dictionary should treat linguistic material “so far as it is *our own*” (Ibid., emphasis mine). What is “our own” can be found in “the general and popular language” (Ibid.: 186) as opposed to professional terminology which “generally derived from other nations” (177). The “foreign” loanwords are acceptable only insofar as they have been “naturalized,” “incorporated” and “considered as part of *our language*” as a consequence of “long intermixture and frequent use” (178, emphasize mine).

Johnson’s worldview is neatly hierarchical: some words are, simply put, better than others, the same way that some writers are “of the first reputation” while others are “of an inferior rank” (191). Selecting quotations from the former should be preferred to those from the latter. The quotations themselves should be selected not only for “their immediate use” but also so that they “may give pleasure

or instruction by conveying some elegance of language, or some precept of prudence, or piety.” (191) Language can never be just language: it is always implicated in value judgements. Yet, despite his unabashed propagation of conservative social values (“All change is of itself an evil, which ought not to be hazarded but for evident advantage” [180]), Johnson is flexible — and realistic — enough to allow for exceptions to his own rules:

Of those [loanwords] which yet continue in the state of aliens, and have made no approaches towards assimilation, some seem necessary to be retained, because the purchasers of the dictionary will expect to find them. Such are many words in the common law, as *capias*, *habeas corpus*, *præmunire*, *nisi prius*: such are some terms of controversial divinity, as *hypostasis*; and of physick, as the names of diseases; and in general all terms which can be found in books not written professedly upon particular arts, or can be supposed necessary to those who do not regularly study them. (178-9)

A dictionary needs users and users come with a set of expectations. Because users may expect to look up loanwords in his dictionaries, and because loanwords are used in various branches of knowledge that should be accessible to the general audience, Johnson felt that his dictionary, being a tool and not simply a theoretical tract, will need to respond to this sociolinguistic fact:

The value of a work must be estimated by its use: it is not enough that a dictionary delights the critic, unless at the same time it instructs the learner; as it is to little purpose, that an engine amuses the philosopher by the subtlety of its mechanism, if it requires so much knowledge in its application, as to be of no advantage to the common workman. (177-178)

Ultimately, however — and in full opposition to the rhetorical modesty with which Johnson opened his text — it is the lexicographer, i.e. Johnson himself, who is the ultimate *arbiter elegantiae* and source of authority on all things linguistic. Even the

best writers can sometimes resort to the use of “barbarous or impure words and expression” (190), but it is the role of the lexicographer — the not-so-harmless drudge, after all — to “review all [the] polite writers and correct such impurities as might be found in them, that their authority might not contribute, at any distant time, to the depravation of the language” (190). Dictionaries may be tools, but lexicographers are not simple tool-makers.

1.2.3.2 *Dictionary as a functional tool*

Throughout the nineteenth and early parts of the twentieth century, lexicology and lexicography enjoyed a more or less symbiotic relationship in which lexicology provided general linguistic models and lexicography — their empirical realizations (Geeraerts 1996). Only in the second half of the 20th century, however, did a number of scholars such as Sčerba (Шчерба 1940), Zgusta (1972), Hausmann (1977), and Kromann *et al.* (1984) start exploring independent theoretical approaches to lexicographic tool-building.

The most ambitious and far-reaching project of conceptualizing and developing a general theory of lexicography in the 20th century was made by Wiegand (1998). According to Wiegand, lexicography is a scientific practice of producing reference works on language, while metalexicography is constituted by four different subdisciplines: the history of lexicography, a general theory of lexicography, research on dictionary use and dictionary criticism (Wiegand 1989). Wiegand defined the dictionary as a utility product (*Gebrauchsgegenstand*) that should satisfy the needs of the user, while developing a detailed and layered typology of language dictionaries (*Sprachwörterbücher*), dictionaries for special purposes (*Sachwörterbücher*) and encyclopedic dictionaries, which treat both linguistic and extra-linguistic matters (*Allwörterbücher*). Above all, his work is a

monument to structuralist thinking: Wiegand defines the purpose of the dictionary phenomenologically, from within, by analyzing the way dictionaries are constructed, both on the micro- and macrolevels.

Der genuine Zweck eines Gebrauchsgegenstandes oder der einer seiner funktionalen Teile wird mithin stets aufgrund einer phänomenologisch basierten Unterscheidung bestimmt. Es geht um die Typen der erhältlichen lexikographischen Information und damit um die Typen der Benutzerfragen. (Wiegand 2001: 230)

[The genuine purpose of a utility object or one of its functional parts is always determined on the basis of phenomenological differentiation. It's a matter of obtainable lexicographic information and with it the types of user questions.]

Wiegand is aware of the potential difficulties caused by the opposition of “theoretical lexicography” to “practical lexicography” because such a terminological distinction implies a clear-cut division of intellectual labor which is neither realistic nor easy to sustain. By separating theory from practice, we run the risk of “attributing reflective components of lexicographic processes to theoretical lexicography, which would leave the compilation technique seem like no more than a vacuous implementation” (Wiegand 1989: 253). Some practicing lexicographers such as Atkins and Rundell, have claimed that “theoretical lexicography” does not exist, but that there is “an enormous body of linguistic theory which has the potential to help lexicographers do their jobs more effectively and with greater confidence” (2008: 4). But not all practicing lexicographers have shared this view. Most notably, proponents of the so-called Modern Function Theory of Lexicography (Bergenholtz and Tarp 2002)¹⁷, which has been under development by a group of

¹⁷ FT did not reach maturity overnight. The historical development of the theory and its initial links to Wiegand's general lexicographic theory are, however, beyond the scope of my project.

researchers associated with the Center for Lexicography at the Aarhus School of Business in Denmark, describe Atkins and Rundell's views as "an expression of British arrogance" and "a threat to the further development of lexicography" (Bergenholtz and Bergenholtz 2011: 189). Even though academic discourse — like any professional discourse — is not entirely immune to personal acridity and bursts of irrationality that give credence to Sayre's law of trivocracy,¹⁸ the virulent hostility with which some proponents of the Modern Function Theory of Lexicography have been attacking the linguistic underpinnings of dictionary making is astonishing — even by the most lax academic standards.

Leaving aside the question of whether these types of attack can and should be attributed to personal eccentricities or the structural dysfunctionality of the academic system as a whole, the rhetoric of "linguists who also masquerade as lexicographers on the side" (Bergenholtz and Bergenholtz 2011: 188), "linguistic colonialism" (Fuentes Olivera and Bergenholtz 2011: 3, 4, 13; Gouws 2011: 22, 27), "linguistic liberation" (Gouws 2011: 25,26) and the utopian announcements of a "a new destiny" (Ibid.: 26) for lexicography point to a great deal of apprehension about the nature of the lexicographic enterprise, its scope as a scholarly discipline, its cultural patrimony and its place in the increasingly technological world.

Unlike Wiegand's structuralist approach whose primary object of analysis is the dictionary itself, albeit as a utility product, Function Theory (hereafter FT) is founded upon the concept of lexicographic functions:

For an interesting chronology that led to the genesis of the new theory, see Chapter 4 in Tarp (2008).

¹⁸ "The reason the politics of academe are so intense is because the stakes are so low" (Homer and Levine 1985: 241)

A lexicographical function is the satisfaction of the specific types of lexicographically relevant need that may arise in a specific type of potential user in a specific type of extralexicographical situation (Tarp 2008: 81).

In other words, FT is a theory “from without.” Its point of departure is not the dictionary as an information-bearing object, but a potential user who finds himself or herself in need of certain information. Accordingly, a dictionary is not a model of language but, above all, a *practical tool* aimed at solving practical problems “which may or may not be related to linguistics” (Nielsen 2011: 198).

FT distinguishes between *primary* or *function-related* user needs, which arise in extra-lexicographic situations (Tarp 2008:41) and *secondary* or *usage-related* user needs, which occur during the dictionary consultation when “an actual dictionary user may need guidance in the form of secondary lexicographical data to find and interpret the primary lexicographical data needed to satisfy their primary user needs” (Tarp 2008: 42).

In terms of the types of situation in which a user may need to consult a lexicographic work, FT distinguishes between cognitive, communicative, operative and interpretive situations. Communicative situations are situations in which users need help solving problems related to acts of communication, whether receptive or productive; cognitive situations are those in which users want to learn more about a certain topic; users in operative situations need help in performing manual or mental operations (in how-to-books, user guides, various indexes etc.); while those in interpretive situations need help understanding non-textual and non-verbal phenomena such as symbols (Tarp 2011: 65). Especially the cognitive situations pose problems because it is “very difficult to select and prepare lexicographical data that

might directly satisfy users' needs in all, almost all, or even the majority of consultations related to cognitive situations" (Ibid.).

FT exhibits a strong preference for *monofunctional* dictionaries, i.e. dictionaries that have only one function, i.e. satisfy only one type of user need. The concern that traditional *polyfunctional* dictionaries offer an overwhelming amount of data which can lead to access problems is a valid and important one. Proponents of the FT are right to insist on the importance of user profiling (in terms of the mother tongue, level of expertise in L2, general knowledge, proficiency in LSP L1 and LSP L2, and type of activity that the user is involved in) for producing dictionaries that correspond to user needs. But the conclusion that "general-language lexicography focuses on the linguistic phenomena instead of on the user's needs" is a hardly justifiable blanket statement. It is enough to think of the tradition of English L2 lexicography and the advances made in producing user-friendly dictionaries geared specifically towards the needs of L2 students (Dolezal and McCreary 1999; Tono 2001; Harwood 2002; Burkhanov 2004; Rundell 2008; Verlinde and Binon 2009).

FT recasts lexicography as a social science which deals with information management, rather than language as such. At the most abstract level, the task of lexicography is to "trace and examine social needs," and, more concretely speaking, to "[classify] the various types of person, situation and need in question, and then [find] out which of these needs can be satisfied using lexicographical data" (Tarp 2008: 41) A question that poses itself at this point would be: how are these needs determined? How does the lexicographer get to know his or her users and their situations of use? And how does he or she then "find out" which of these needs can be satisfied using lexicographic data? FT requires that the user needs be established before embarking on a lexicographic project, but it does not actually propose any

type of empirical research into user needs or situations of use. On the contrary, Tarp is explicitly rejecting the possibility of empirical research:

But how can theoretical lexicography find the relevant situations? In principle, it could go out and study all the hypothetical social situations in which people are involved. But that would be like trying to fill the leaking jar of the Danaids. Instead, initially lexicography needs to use a deductive procedure and focus on the needs that dictionaries have sought to satisfy until now, and on the situations in which these needs may arise. (Tarp 2008:44)

It is not clear why Tarp rejects empirical research, when this kind of research into dictionary use has already been quite fruitful since the late nineties (see, for instance Kosem *et al.* 2018; Müller-Spitzer *et al.* 2015; Müller-Spitzer 2014; Rundell 1999; Atkins 1998; Atkins and Varantola 1997). Furthermore, by suggesting that dictionary makers should simply analyze the kind of needs that the existing dictionaries, i.e. those developed before FT, have sought to satisfy, Tarp is committing the same “error” that he accuses Wiegand of:

Such a procedure is typical for the theoretical approach that first of all contemplates and interprets already existing dictionaries. (Bergenholtz and Tarp 2003: 187)

The decision what to take as a starting point for one’s theory of lexicography — lexicographic functions, or structural and data-centric considerations — is not a matter of right or wrong, but a matter of choice, which should be recognized as such. I do not dispute the importance of user needs or situations of use. Nor do I find any fault with FT’s claim that the dictionary is a tool whose primary function is to help users solve concrete problems. This chapter and this dissertation as a whole argue that a theory of lexicography that focuses *only* on one particular facet of the

dictionary (the dictionary as a tool) while completely ignoring or negating the other facets of the dictionary (as a text, a model of language and an artifact of culture) oversimplifies the multiple roles — and functions — that the dictionary can have.

Even when considering the instrumental aspect of the dictionary, FT reduces its function to direct and minimal information extraction:

User needs are always needs for information which can be extracted from lexicographical data. Once this information has been extracted, it can be used to solve the communicative problems that led to the dictionary consultation in question. But it can also take root in the brain in the form of knowledge, even though this was not the purpose of the dictionary consultation. Such knowledge is quite simply a bonus which is in principle irrelevant. But in some types of dictionary — learner's dictionaries, for instance — this »bonus« may actually be extremely relevant. (Tarp 2008: 87)

While it is entirely plausible to claim that the *primary* function of a certain type of dictionaries is to provide quick and easy access to information which solves particular communication problems — or, if we were to put it slightly more cautiously: that certain users may consider target information extraction as the sole purpose of the dictionary — it is much less convincing to claim that knowledge gained while consulting dictionaries is “in principle irrelevant.” If that were the case, and the knowledge “bonus” was relevant only to the consultation of learner's dictionaries, but no other type of dictionary, that would also mean that no user will ever consider consulting any other type of dictionary with the explicit purpose of gaining knowledge. Or, at the very least, that FT cannot imagine the existence of a user, either a scholar or a language enthusiast, who consciously decides to pick up a general English-Danish bilingual dictionary (and not one which is explicitly conceived, compiled and/or marketed as a learner's dictionary) because he or she

wants to learn (and remember) how to say, for instance, *nonsense* in Danish. Or a user who wants to consult a polyfunctional, academic dictionary like the *OED* in order to learn and, perhaps, even not forget, the year of the first citation of the word *nonsense* in the English language.¹⁹

By focusing on the simplified and rigid situations of use and downplaying the role that different types of data play in the consultation process, Tarp ends up making an exaggerated and unsubstantiated claim that data structures are not relevant in the consideration of the role that they play in a given dictionary :

it is necessary to underline that the category to which specific data belongs says nothing (and can say nothing) about the specific purpose of this data in a given dictionary, since *this purpose depends on the function or functions that this dictionary is designed to have*. (Ibid.: 100, emphasis mine)

Even though FT approaches dictionary functions from the extra-linguistic point of view of a potential user, rather than from the structure of the dictionary itself, it is extravagant, to say the least, to claim that the type of data in a dictionary “can say nothing” about its purpose. The specific purpose of any structural element in a given dictionary is to provide information about a certain linguistic or extra-linguistic phenomenon: whether it is an indication of a part of speech, a semantic domain, or a stylistic or pragmatic label, each constitutive element of a dictionary entry has its own purpose *by design*. A purposeless structural element in a dictionary would be difficult to imagine.

FT reduces the scope of what it considers useful dictionaries by placing emphasis on monofunctional as opposed to polyfunctional dictionaries, while, at

¹⁹ “1612. A. Stafford *Meditations & Resolutions* 101 Others againe, by an vnwittie application, make *non-Sense*; and infuse lothing into the nice stomach of the Reader. I wil apply nothing, which may not imply something.” (“nonsense” – OED)

the same time, extending the subject of the theory of lexicography to include the production of manuals, how-to-books, user guides, indexes etc.²⁰ This kind of approach points to an inherent tension in the disciplinary struggle at the core of the FT enterprise: to reduce the role of linguistics in lexicography as a way of securing for it the independent status of a science:

the planning and compilation of certain dictionaries have no need for any linguistic input. (Gouws 2011: 25)

While this may, arguably and to some extent, be true in the case of LSP dictionaries, it is hardly true for LGP dictionaries. The imprint of LSP on FT is not to be underestimated (see for instance Tarp 2000). The complete disregard for linguistics makes FT more appropriate as a theory of LSP dictionaries.

In the following section, I will turn to one particular function of the dictionary which plays no role in Function Theory but which is essential for my digital edition of Vuk Stefanović Karadžić's *Lexicon Serbico-Germanico-Latinum*: the dictionary as a research tool.

1.2.3.3 *Dictionary as a research tool*

All dictionaries satisfy some information needs, but historical — non-contemporary — dictionaries are also research objects in their own right. It is highly unlikely that speakers or students of English will nowadays resort to Johnson's *Dictionary of the English Language* (1755) for their day-to-day information needs. The same goes for Adelung's *Grammatisch-kritisches Wörterbuch der hochdeutschen Mundart*

²⁰ "A general theory of lexicography will focus on lexicographic work, but should not isolate itself from the theory underlying the development, planning, compilation and publication of other reference sources" (Gouws 2011: 26). "Moreover, the list can easily be extended, to include even telephone books, internet-based search engines and other similar reference tools" (Tarp 2011: 56).

(1774–1786), Dal’s *Толковый словарь живого великорусского языка* (1863–1866) or any other similar historical dictionary. Historical dictionaries, however, are important, and often fascinating, objects of study because, among other things, they reveal sociolinguistic perceptions and culturally-shaded conceptualizations of a particular epoch’s lexical knowledge.

Function Theory does not concern itself with historical dictionaries at all. Yet historical dictionaries are still lexicographic tools who have their users, and those users have their lexicographic needs. It has already been observed, outside the context of Function Theory, that users resort to historical dictionaries in roughly three different types of situations: (1) when they have difficulties in the reception of historical texts, (2) when they have difficulties in the production of modern translations; and (3) when they have general questions about linguistic and cultural traditions (see Reichmann 2012: 54). The first two types of situations are text-related: they arise out of the user’s engagement with a particular text. The user can, when reading historical texts, experience all sorts of semantic difficulties (encounter unknown lexical units; discover gaps in word meaning; raise questions of morphological, syntactic or pragmatic nature). But because these questions are related to a particular text and a particular linguistic aspect of it, the use of the historical dictionary will in this case not differ from the consultation of any other type of dictionary. The user will use the macrostructure (or the search engine, in case of an e-dictionary) to locate a specific entry containing information he or she needs.

Reichmann’s third situation of use is *texttranszendierend* [text transcending] (ibid.: 64). What this means is that dictionaries can be used in the study of the way historical relations, processes and transformations register in language. Historical

dictionaries provide valuable information about how words change their meaning through time. Consequently, lexicographic text analysis can be seen as a tool for historical semantics (see, for instance Lobenstein-Reichmann 1998; Lobenstein-Reichmann 2000). The main difference between the use of dictionaries in specific text-reception and text-production situations, on the one hand, and more general research situations is the question of initial focus and ultimate scope. In specific, text-related situations of use, the initial focus and ultimate scope are usually the same: extracting the definition of a particular sense of a particular word is usually accomplished by consulting one dictionary entry. In research situations, the initial focus is really just a point of departure for broader considerations: for instance, the study of labor relations as represented in nineteenth-century Serbian dictionaries could not be accomplished by consulting one entry — *pađ* [work, labor] — but would require the study of an entire semantic field related to the concept of work; its synonyms and antonyms; its hypernyms and hyponyms, its collocates etc. While the initial focus of a text-transcending study may be a single dictionary entry, its ultimate scope is a semantic network of entries. In text-specific situations, the dictionary is used as a look-up tool. In text-transcending situations, it is used as an exploratory tool.

The notion of the historical dictionary as an exploratory tool is central to the digital edition that forms an integral part of this dissertation. In Chapter 3 *Data modeling as an interpretative framework*, I will argue that the structural and semantic encoding of a historic dictionary can be used as both a model of the text and a model for a digital edition. And in Chapter 6, *Implementing an API-centric digital edition*, I will show how an API-centric digital edition of *LSGL* was built around the TEI-encoded dictionary text in order to serve as an exploratory platform.

1.2.3.4 *Dictionary as a digital tool*

The discussions of the dictionary as an ideological, functional and a research tool above have focused on these three aspects of the dictionary regardless of the medium in which the dictionary is compiled or delivered to the user. This is because the dictionary — any dictionary — can serve an explicit ideological message or echo an implicit ideological purpose, while being functionally useful to its users and potentially an object of research in its own right, regardless of whether it is a print dictionary or an electronic dictionary. This does not mean, however, that the role of the medium is nonessential, especially not in our considerations of what the dictionary as a functional or exploratory tool lets its users do with it. The rest of this section will explain why.

An electronic (or online)²¹ dictionary should be regarded as “a complex whole in the two-dimensional space provided by the image on a (computer) screen with data lists and outer texts, referred to as surface features” as well as three underlying features: (1) the dictionary is designed to fulfill one or more functions; (2) the dictionary contains data that has been selected with the goal of fulfilling its functions; and (3) the dictionary has structures and links that marshal its data into fulfilling its functions (Nielsen 2011: 198). In other words, online dictionaries are “in effect search engines that search for structured data in a database, retrieve the relevant data, and present them to users in predetermined ways” (Nielsen and Almind 2011: 166).

As a type of search and retrieval tool, an online dictionary consists of: (1) a

²¹ The distinction between electronic or e-dictionaries and online dictionaries is no longer maintainable since dictionaries are, as a rule, no longer delivered on optical memory devices (CDs and DVDs). For all intents and purposes, e-dictionaries and online dictionaries are nowadays fully synonymous terms.

structured database which stores lexicographic data in appropriate fields (for instance: lemmas, grammatical information, definitions) and provides a relational mechanism for connecting these separate fields; (2) a surface view or what the user sees on the screen: a selection of information from the database; and (3) a search-engine which functions as the intermediary between the database and the dictionary (ibid.: 147). The most important consequence of this three-partite structural understanding of an electronic dictionary is simple but far-reaching: one database can contain, constitute and produce multiple dictionaries.

Function Theory takes the concept of “one database, multiple dictionaries” to the extreme. Andersen and Almind describe “three different dictionaries” of English phrasal verbs: one which assists dictionary users in the reception of English-language texts with phrasal verbs; one which helps with the translation of English-language texts with phrasal verbs into Danish; and one that assists in the production of English-language texts with phrasal verbs (Andersen and Almind 2011: 210). In a similar vein, Bergenholtz speaks of four different dictionaries based on a database of Danish fixed expressions, only to conclude that with advanced search options “each user will be able to design one’s own individual dictionaries. It will therefore be possible to extract not only four, but thousands of different dictionaries from the database” (Bergenholtz 2011: 53). All those different “dictionaries”, however, are, in fact, only different, customized views of the same dictionary, which, depending on the purpose, reveal or hide some of the dictionary components from the user. Different views or different slices of the data saved in the database can play an important role in providing useful functionalities for the user of an online dictionary, as I show in Chapter 6 *Implementing an API-centric digital edition*, but awarding quasi-ontological status to each different view is neither convincing nor

necessary. It reduces “the dictionary” to the data contained within it. It decontextualizes it as a product of humanistic scholarship *and* as a historical artifact. And, at the same time, it uncovers a contradiction within Function Theory as such: the “thousand-dictionary-claim” actually highlights the continued value of polyfunctional dictionaries. If one set of lexical data can produce multiple views, it only speaks to the quality, density and complexity of that data.

Setting aside the excessive multiplication of “dictionaries” that underpin the database-sourced lexicographic works described by FT, the technical implementation of the dictionary as a search engine significantly effects the way we access dictionary content. The alphabetical arrangement of entries in a print dictionary functions as a type of *index* — a retrieval mechanism connecting a known order of symbols to an unknown order of information (Hass Weinberg 2010). The user can access dictionary content by consulting the dictionary macrostructure, i.e. the arrangement of lemmas in a given order (see Hausmann and Wiegand 1989). While alphabetic dictionaries are relatively easy to consult, they are also efficient randomizers of meaning.²² By grouping lexemes according to their orthography, rather than their sense, standard dictionaries adhere to the abstract convention of alphabetical order, scattering words with similar or related meaning across unpredictable distances. The “psychologically quite unmotivated tyranny of the alphabet” (Makkai 1980: 127) is both a blessing and a curse. Looking up entries is easy, if one knows precisely what word one is looking for. Discovering unfamiliar

²² For a useful overview of different types of organizing Chinese dictionaries, both graphological (using radicals and the number of strokes) and phonological (based on sound), see Saarela (2014).

words and exploring semantic concepts, however, is considerably more difficult (Tasovac 2012).

In electronic dictionaries, on the other hand, users access lexicographic content not based on a single wordlist but through a search engine: “it may be more appropriate to say that the macrostructure has been replaced by what may be called a data presentation structure.” (Nielsen 2011: 201; see also Nielsen and Almind 2011). The lexicographic concept of accessibility needs to be “narrowed down to cover *quick and easy access* to the specific types of data that can cover a specific type of user’s specific types of need in a specific type of extra-lexicographical situation” (Tarp 2008: 101).

The proponents of Function Theory take a certain amount of pride in the fact that their approach, unlike Wiegand’s, is not medium-specific:

When a given function determines the way in which a given dictionary is presented, it does not matter whether it is a printed or an electronic dictionary. The theory requires no adaptation to embrace all different types of dictionaries within its scope. (Gouws 2011: 24)

This is a kind of claim that can be made only at a very high level of abstraction. The question remains, however, whether choosing a vantage point high enough to be able to gloss over the differences between particular phenomena and their media-specific realizations makes a theory not only less practical but also less useful.

As far as electronic lexicography is concerned, the greatest potential of FT seems to be, indeed, in the area of individualization and user profiling. While the idea of filtering content based on user needs is highly pertinent — and ideally suited for software environments that make such content negotiation possible — difficulties can arise at two different levels: in terms practicality for each individual

user, and in terms of manageability across various electronic dictionaries.

Bergenholtz and Nielsen (2006) list eight questions that are important for drawing up a user profile. Nielsen and Almind (2011) adapt that list for their *Accounting Dictionaries* and present a diagnostic checklist containing ten questions:²³

It remains to be seen whether the simplicity of using monofunctional tools will not be overshadowed by the complexities of setting up detailed user profiles for each such tool. Furthermore, user profiling may create an unrealistically static image of the user. Use scenarios change. In one consultation, the user might be interested in morphology; next time in pronunciation, and in yet another in semantics. Having to change one's profile for each query in a polyfunctional dictionary in order to satisfy the requirements of monofunctional views would not be necessarily embraced by every user. If, on the other hand, the Internet became a resource consisting exclusively of monofunctional dictionaries, finding and locating an appropriate dictionary for simple look-up tasks would introduce another level of complexity that could make the consultation process much more difficult and time-consuming than it is in a polyfunctional dictionary.

In fact, a preliminary usability study²⁴ of four electronic dictionaries has already shown that many presuppositions of the FT do not necessarily match with actual user expectations. Even though the test persons did consider efficiency to be the most prominent advantage of electronic dictionaries and quick and easy search

²³ What language is their native language? At what level do they master their native language? At what level do they master a foreign language? How extensive is their experience in translating between the language in question? What is the level of their general cultural and factual knowledge? etc.

²⁴ “commonly seen as a cluster of properties of (software) products which ensures effective and efficient use of the software, as well as user satisfaction” (Heid 2011: 287).

functions their most important functionality, they also thought that “the presentation of large amounts of lexicographic data was an advantage in itself” (ibid.: 296). Heid speculates that “the search habits of students are massively influenced by internet search engines, for which they accept being confronted with large amounts of candidate results from which they have to manually select” (296). While this may be the case, there is no reason to suspect that the Internet is going away any time soon, or that the existing internet search engines will dramatically alter their interfaces and overall approaches to delivering information overnight. The students considered “a broad and detailed set of lexicographic indications” to be the most important feature of a good electronic dictionary followed by “comfortable and error tolerant search functions” (296) and “relevance ranking” (297).

Only ten percent of students in this study managed to complete the search for a multi-word expression *c’est une question de vie ou de morte* in the *Base lexicale du français* (BLF) — an electronic dictionary whose unorthodox interface is built with FT in mind as a collection of various small, monofunctional dictionaries instead one, unified search field (Verlinde 2010). More than fifty percent did not find a German equivalent of the French word *cloture*, and only one sixth of the test subjects accessed the translation without difficulties (Heid 2011: 300-1). In the post-test questionnaire, the BLF was rated poorly in terms of access to relevant data and conformity to user expectation (301). Overall, the test results suggest that users are:

reticent to follow deeper search paths where a sequence of decisions has to be taken, for example, going first to a lemma, then to its reading, and finally to its collocations. While such search paths are standard for printed (specialized) dictionaries, they seem to be less efficient (and less well accepted) in electronic dictionaries. (ibid.: 303)

More usability studies such as this will be needed to test and evaluate the application of FT to electronic lexicography.

FT is considered to be a theory of both printed and electronic dictionaries. It makes this claim because it focuses on data extraction in individual use situations. But, as we have already seen above, the electronic dictionary has arguably lost its traditional macrostructure: because users access lexicographic data in electronic dictionaries not based on a single wordlist but based on a search engine, the electronic dictionary is — or has been so far — quite a different text-bearing object than its predecessor. Accessing individual bits of data in order to satisfy basic user needs may be the same type of mechanism that occurs during the consultation of both print and electronic resources, but the experience of the dictionary as a material object let alone cultural symbol could hardly be any more different.

The question here, however, is not only of symbolic nature. The lack of a visible, manageable and finite macrostructure and database-centric design paradigms can have far-reaching consequences for the way we use electronic dictionaries: for the type of browsing we can conduct in them, and for the type of serendipitous discovery that they make possible. I will return to this topic in Section 1.3 *Affordances: the dictionary's actionable properties* (pp. 69-83). At the moment, FT has no interest in serendipitous discovery because it considers it irrelevant from the point of view of targeted information extraction. In the long term, however, this kind of approach can only undermine the possibility of creating new and innovative lexicographic tools.

From a broader perspective, however, the deficiencies of FT are symptomatic of a particular kind of approach to technology. As Willard McCarty points out, the early stages of a new technology tend to be evaluated in terms of

quantity, not quality. Electronic dictionaries are no exception in this respect.

Efficiency, however, should not be the ultimate horizon of any technology:

Tools are... perceptual agents. A new tool is not just a bigger lever and a more secure fulcrum, but rather a new way of conceptualizing the world.
(McCarty 1996)

A truly modern theory of electronic lexicography will have to tackle a series of questions that FT in its current form has no answer for or no interest in: How can electronic lexicography serve society — not just as a sum of individuals with individual needs for information but society as a collective with “nationale, staatliche, politische, administrative oder wissenschaftliche Dokumentationswünsche” [wishes for national, state, political, administrative or scientific documentation] (Hausmann 1989a: 5)? What kinds of interfaces can we and should we build in order to provide access to dictionaries as part of our cultural heritage? And, finally — and most importantly — how can we create tools that, while being user-friendly, are capable of representing complexity both in terms of their lexical architecture *and* their cultural capital? A theory that without any sense of irony lumps dictionaries and phone books in the same category of tools cannot provide those answers.

1.2.4 *Dictionary as a cultural artifact*

No dictionary emerges out of a vacuum. Instead, as we already saw in the case of Johnson’s use of his *magnum opus* as a programmatic tool, a dictionary is always shaped and influenced by the social values of its time. The motors of those values include religion, literature, education, politics, economics and language planning (Hausmann 1989a; Mackintosh 2006). Cultural and ideological perspectives are reflected in the selection of lexemes; the selection of geographic and social idiolects;

the treatment of lexical variations; the value judgements implied in dictionary labels (colloquial, vulgar, jocular); the cultural presuppositions of translatability etc. That is why a dictionary is “a mirror of its time, a document to be understood in sociolinguistic terms” (Kahane and Kahane 1992: 20).

At the same time, dictionaries are not passive receptors of social influence: they are both culturally constructed and culture-constructing (Fishman 1995). This is why dictionaries play such a crucial role in all four stages of language standardization (selection, codification, acceptance and elaboration of a standard — see Haugen 1966).²⁵ While national dictionary projects started emerging in Western Europe from the seventeenth century onward, the continent saw its golden age of vernacularized, nation-building lexicography in the late eighteenth and the nineteenth centuries (for an overview, see Seton-Watson 1977). The Russian Academy produced a six-volume dictionary of the modern, vernacular Russian between 1789 and 1794.²⁶ Vuk Stefanović Karadžić published the first edition of his *Serbian Dictionary* in 1818, while Joseph Jungmann published his influential five-volume Czech-German Dictionary between 1835 and 1839. Long before Webster published his *Compendious Dictionary of the English Language* (1806) and his famous *American Dictionary of the English Language* in 1828, he outlined his plan to reform American spelling in order to “make a difference between the English orthography

²⁵ Zgusta (1989) identifies four types of dictionaries in relation to standards: standard-creating dictionaries, modernizing dictionaries, antiquating (or archaizing) dictionaries, and standard descriptive dictionaries (which try to describe and clarify the existing standard). Vuk Stefanović Karadžić’s *Serbian Dictionary* was a standard-creating dictionary at the time it was published. Basing a modern Serbian standard on this dictionary today would, however, turn it into an archaizing dictionary. Dictionary typologies — like dictionaries themselves — do not transcend time.

²⁶ For an annotated bibliography of Slavic dictionaries up to 1850, see Stankiewicz (1984).

and the American” in a move which would be of “vast political consequence” because “a national language is a band of national union” (Webster 1789/1991: 87).

This type of double bind — the dictionary as both an echo and a shaper of its time — has important consequences for the dictionary as a research object and for our consideration of dictionaries in the digital environment. Treating lexicography as part of cultural heritage, frees “the history of dictionaries from too heavy a dependence on certain potentially arid kinds of narrative of the form ‘61 per cent of the entries in Y derive from entries in X’, and engages it with broader and more humane questions about lexicology, the history of linguistics, the history of learned culture, indeed the history of culture in general” (Considine 2008: 314).

Within the history of culture in general, the dictionary is an embodiment of a particular type of architectonic thinking. The structure of the dictionary and its history of use impose the notion of imaginary finitude upon language while creating a possibility for the infinite combinability of its constitutive parts. The dictionary as *a symbol of infinite finitude and finite infinity* is a fascinating object of cultural reflection.

Paradoxically or, perhaps, as a consequence of the way in which dictionaries make something infinite (language) seem finite, and something finite (dictionary) – infinite, there are two types of cultural logic functioning in the dictionary as a discursive field: the logic of order and the logic of madness or dreams.²⁷ The former is based on the navigability of the system, the findability of its constitutive parts,

²⁷ Madness and dreams are culturally loaded concepts with an elaborate history of use (Freud, Lacan, Foucault etc.) It is unfortunately beyond the scope of this research to analyze different ways in which literary and philosophical appropriations of the lexicographic order reflect upon and critique the logic of human cognition. I mention it here, however, because this train of thought — even though I have so far sketched it only in the most rudimentary fashion — has and will continue to inform my approach to the dictionary as a complex and occasionally contradictory artifact.

the predictability of the tool and the systematicity of the underlying view of language and meaning; the latter on the utter arbitrariness of that systematicity. That is why the dictionary form can be used both to fix and undermine meanings, solidify certain social practices and linguistic patterns as well as reveal their randomness.

On the one hand, we have the typical example of somebody like Samuel Johnson who saw in his *Dictionary of the English Language* an attempt to counter “the tyranny of time and fashion, [...] the corruptions of ignorance, and caprices of innovation” (Johnson 1785: 7). On the other hand, the authors of a whole series of fictionaries — works of humorous lexicography — dictionaries of clichés, missing words, superfluous neologisms or eccentric definitions — have used the dictionary format to deconstruct its transparency and objectivity by stressing the provisional nature of social constructs as reflected in dominant language use.²⁸

²⁸ These include Bierce’s *The Devil’s Dictionary* (originally published as *The Cynic’s Word Book*) (1906/2000); Flaubert’s *Dictionnaire des idées reçues* (1913/2002); Leiris’ *Glossaire: j’y serre mes gloses* (1925a; 1925b; 1925c; 1939; 1969) ; Finkelkraut’s *Petit fictionnaire illustré: Les mots qui manquent au dico* (1981); Douglas Adams’ and John Lloyd’s *The Meaning of Liff* (1984) and *The Deeper Meaning of Liff* (1990), Desproges’ *Dictionnaire superflu à l’usage de l’élite et des bien nantis* (1985); Kelly-Bootle’s *The Computer Contradictionary* (1995); Ross’ *My First Dictionary: Corrupting Young Minds One Word at a Time* (2011); Lefler’s *The Chicktionary: From a-Line to Z-Snap, the Words Every Woman Should Know* (2011), to name just a few. Electronic resources in this category include the crowdsourced website *Urban Dictionary* (<http://urbandictionary.com>) which started as a dictionary of slang but over time evolved into a popular platform for humorous lexicography; *Pseudodictionary: The Dictionary for Words that Wouldn’t Make It into a Real Dictionary* (<http://pseudodictionary.com>); the Serbian counterpart to the Urban Dictionary, *Vukajlija* (<http://vukajlija.com>) whose name is a pun on both Vuk Stefanović Karadžić, whose dictionary is the subject of my dissertation, and the author of a popular lexicon of foreign words, Milan Vujaklija (Вујаклија 1937/2004); the Italian *Sveltopedia* (<http://sveltopedia.com>) which imposes an additional constraint on its users to define terms via Twitter in less than 140 characters each (“Ikea: Tecnica ipnotica per costringere la vittima a mangiare aringhe crude con salse dai colori sgargianti.”) etc. For an interesting collection of examples of texts that quote or mimic dictionary entries in advertising and popular culture, see Nagy (2004).

The paradoxical cultural value of the dictionary is perhaps best illustrated by the multi-authored *Dictionnaire critique* which was published as a regular feature in the short-lived art magazine, *Documents* (1929-1930) and edited by Georges Bataille (see Noys 2000). Bataille's contributions to the dictionary with entries such as *Abattoir*, *Architecture*, *Bouche* [mouth], *Cheminée d'usine* [factory chimney], *Œil* [eye], *Musée* [museum] and *Poussière* [dust], often in juxtaposition with striking images, emphasized the peripheral, the associative and the contingent rather than the traditional lexicographic meaning of the terms. In his entry on the factory chimney, Bataille recounts, for instance, how these “épouvantails géants” [giant scarecrows] (1929b: 329) instilled fear and anguish in him as a child “comme la révélation d'un état de choses violent dans lequel il se trouve pris à partie” [as the revelation of a state of violence for which one bears some responsibility]:

A cette manière de voir enfantine ou *sauvage* a été substituée une manière de voir savante qui permet de prendre une cheminée d'usine pour une *construction de pierre formant un tuyau destiné à l'évacuation à grande hauteur des fumées*, c'est-à-dire pour une abstraction. Or, le seul sens que peut avoir le dictionnaire ici publié est précisément de montrer l'erreur des définitions de ce genre. (Bataille 1929b: 332)

[This childish or untutored way of seeing is replaced by a knowing vision which allows one to take a factory chimney for a stone construction forming a pipe for the evacuation of smoke high into the air — which is to say, for an abstraction. Now, the only possible reason for the present dictionary is precisely to demonstrate the error of that sort of definition.]

The dictionary, according to Bataille, represents a manner of seeing the world (“une manière de voir”); it imposes a framework of socialized, civilized knowledge (“une vision savante”) upon the more primitive, instinctual and unmediated experience of reality (“une vision sauvage”). While Bataille's essays can be understood only in

terms of his wider exploration of materialism, violence and transgression, his employment of the dictionary format for the denunciation of the kind of knowledge that the very format represents seems emblematic of the cultural double-bind that surrounds dictionaries in our society.

Take, for instance, Bataille's entry *Informe* [formless], which suggests a dictionary should not be concerned with meanings at all:

Un dictionnaire commencerait à partir du moment où il ne donnerait plus le sens mais les besognes des mots. Ainsi *informe* n'est pas seulement un adjectif ayant tel sens mais un terme servant à déclasser, exigeant généralement que chaque chose ait sa forme. Ce qu'il désigne n'a ses droits dans aucun sens et se fait écraser partout comme une araignée ou un ver de terre. Il faudrait en effet, pour que les hommes académiques soient contents, que l'univers prenne forme. La philosophie entière n'a pas d'autre but: il s'agit de donner une redingote à ce qui est, une redingote mathématique. Par contre affirmer que l'univers ne ressemble à rien et n'est qu'*informe* revient à dire que l'univers est quelque chose comme une araignée ou un crachat. (Bataille 1929a: 382)

[A dictionary would begin when it no longer gave the meaning of words, but their tasks. Thus *formless* is not only an adjective having a given meaning, but a term that serves to downgrade, generally requiring that each thing have its form. What it designates has no rights in any sense and gets itself squashed everywhere, like a spider or an earthworm. In fact, for academic men to be happy, the universe would have to take shape. All of philosophy has no other goal: it is a matter of giving a frock coat to what is, a mathematical frock coat. On the other hand, affirming that the universe resembles nothing and is only formless amounts to saying that the universe is something like a spider or spit.]

The conceptual implications of the adjective *formless/shapeless* allow Bataille to “think the removal of all those boundaries by which concepts organize reality, dividing it up into little packages of sense” (Krauss 1985: 39). By criticizing traditional lexicographic definitions for their abstraction, for the fact that that they

turn images into signs which are decontextualized from both the psychological and social aspects of reality, Bataille reappropriates the dictionary genre — a traditional instrument of culture — as a tool of symbolic, cultural transgression. *Dictionnaire critique* is a dictionary about the insufficiency of dictionaries but also about their malleability. Bataille criticizes lexicographers — and philosophers, for that matter — for giving shape to what is, in reality, shapeless. But he also chooses to cast his critique of traditional dictionaries and language-based knowledge systems in the shape of a dictionary itself. In its most abstract sense, I would claim, the dictionary is much more than an instantiation of a particular linguistic worldview. It is an articulation and a projection of the *possibility* of meaning. No dictionary can truly guarantee anything else but to embody this possibility. In the end, the very *idea* of the dictionary makes it possible to create an anti-dictionary like Bataille's *Dictionnaire critique*. The catch is that an anti-dictionary still remains a dictionary.

1.3 Affordances: the dictionary's actionable properties

In Section 1.2.3 *Dictionary as a tool* (pp. 39-63), I have already outlined my criticism of the Function Theory of Lexicography, both in general and as it pertains to historical dictionaries in particular. Even though FT focuses on the user and the user's needs, it fails to account for the situations of use involving historical dictionaries. In addition, FT does not address the subtle differences in the kind of user actions and information-seeking behaviors that are possible in the print and the digital medium. In the following sections, I will remedy the deficiencies of the Function Theory by conceptualizing dictionary *affordances* and by showing how different access paths,

which involve both the structure and the presentation layers of print and electronic dictionaries, affect what users can do with a dictionary.

1.3.1 *The concept of affordance*

Gibson (1979) introduced the concept of *affordance* in the context of an ‘ecological’ approach to visual perception as “what [the environment] *offers* [...], what it *provides* or *furnishes* [...]” (127). According to Gibson, when we perceive objects, we do not discriminate their physical properties or qualities but rather their potential uses. The term itself, Gibson claims, is original: “The verb *to afford* is found in the dictionary, the noun *affordance* is not. I have made it up.” (127) Gibson’s claim to lexical innovation is partly justified: while the use of *affordance* to describe the readily perceivable utility of an object is indeed his contribution to the English language, the lexeme itself was recorded as an English dialectism at the end of the 19th century in the sense of “ability to bear expense” (Dickinson [1879] as quoted by the OED). The notion of objects inviting certain kinds of actionable behavior has a rich history, which Gibson himself traces to Koffka’s “demand character” as outlined in his *Principles of Gestalt Psychology* (Koffka 1935) and Lewin’s notion of *Aufforderungscharakter* from his *Grundzüge der topologischen Psychologie* (Lewin 1969). The main difference between a Gestaltist notion of *valences* and Gibson’s affordances lies in the fact that the latter are invariant and not dependent on the observer’s needs or his act of observing: “The object offers what it does because it is what it is” (Gibson 1979: 139).

Ecological psychologists like Gibson reject the view that animals and humans construct rich mental representations of objects out of basic stimuli (points of light, wavelengths etc.), arguing instead that both animals and people perceive directly the information that is relevant to their needs. In other words, what we see

is not so much objects themselves, but rather the acts or behaviors that are *afforded* or made possible by the objects:

We would say that humans do not perceive chairs, pencils, and doughnuts; they perceive places to sit, objects with which to write, and things to eat. (Michaels and Carello 1981: 42)

In the context of human-computer interaction, Norman (1988) argued that operating any type of tool successfully depends on three requirements: the tool's conceptual model, its constraints (physical, logical and cultural) and its affordances or, rather, perceived affordances, as he stresses in a later article (1999: 39).

According to Norman, good design shapes artifacts in such a way that its possible affordances may be easily recognized by its users. Because screen-based objects come with their own set of physical, built-in affordances of the screen itself, the the designer can really only work with perceived affordances:

designers sometimes will say that when they put an icon, cursor, or other target on the screen, they have added an “affordance” to the system. This is a misuse of the concept. The affordance exists independently of what is visible on the screen. Those displays are not affordances; they are visual feedback that advertise the affordances: they are the perceived affordances. The difference is important because they are independent design concepts: the affordances, the feedback, and the perceived affordances can all be manipulated independently of one another (Ibid.)

In the context of this research, I am not exploring the dictionary interface in terms of design or interface aesthetics (even though aesthetic factors and issues of functionality are undoubtedly mutually dependent; see, for instance Dillon 2001 and Petersen *et al.* 2004) but precisely in terms of affordances — the kinds of information-seeking behaviors that dictionaries invite, provoke and encourage. Affordances are themselves *functional* in the sense that they are “enabling, as well as

constraining factors” (Hutchby 2001: 448). As such they should be seen as the underlying principles that make certain design choices (and, consequently user actions) possible. It does not make sense to discuss navigational structures, buttons or link colors in an online dictionary *before* we have established the affordances of the dictionary as preconditions for the user’s interaction with it. In the following sections, I will discuss the way in which the dictionary offers opportunities for action, and, especially, how these opportunities change when we move from the printed page to the computer screen.

1.3.2 *Functional affordances*

More recently, the concept of affordances has been recognized as a productive one in the field of information systems as well, especially as a useful device for establishing middle ground between technological determinism and social constructivism (see, for instance, Faraj and Azad 2012; Leonardi and Barley 2010; Robey *et al.* 2013; Stendal *et al.* 2016). In the context of my discussion of the dictionary as a complex socio-technical assemblage which affords certain kinds of information-seeking behavior while holding certain symbolic, interpretative values, the notion of *functional affordances* as developed by Markus and Silver (2008) proves to be particularly useful:

Functional affordances are a type of relationship between a technical object and a specified user (or user group) that identifies what the user may be able to do with the object, given the user’s capabilities and goals. More formally, *functional affordances* are defined as the possibilities for goal-oriented action afforded to specified user groups by technical objects. (Ibid.: 622)

While Markus and Silver focus on information systems in general, I use the notion of functional affordances as possibilities for information-seeking actions allowed by

the dictionary as an assemblage, regardless of its medium (analogue or digital), or to put it more simply, as that which a user might be able to do *with* a dictionary, given the user's capabilities and goals.

1.3.3 Types of information-related behavior

Bates (2002) identified two distinct sets of variables that determine the mode of information-related behaviors: (1) active vs. passive, and (2) directed vs. undirected. By cross-tabulating the variables, Bates arrived at four distinct modes of information seeking behavior, which she conceptualized as: (a) searching, (b) browsing, (c) monitoring and (d) being aware.

Table 1: Modes of information-related behavior according to Bates (2002)

VARIABLES	ACTIVE	PASSIVE
DIRECTED	(A) Searching	(C) Monitoring
UNDIRECTED	(B) Browsing	(D) Being Aware

Searching and browsing satisfy *recognized* but *not yet fulfilled* needs, whereas monitoring and being aware meet and fulfill *unrecognized* information needs, i.e. the kind of needs that are recognized as such only upon being fulfilled. *Searching* and *browsing* are both active because the user is taking concrete steps in pursuit of information, although the information sought is of different scope. In the case of *searching*, the user is looking for specific information; in the case of *browsing*, the user does not necessarily have particular information needs, but recognizes the usefulness of actively exposing himself or herself to information. When it comes to passive information-related modes, *monitoring* describes “a back-of-the-mind alertness for things that interest us” (5) without a pressing need to engage in active pursuit of that particular information. This is similar to what Wilson (1997) called

passive search which is to say when “one type of search (or other behavior) results in the acquisition of information that happens to be relevant to the individual” (562). Finally, what Bates calls *being aware* is a type of information absorption that takes place in various social situations whose explicit purpose is not acquisition of information or the kind of information that is absorbed from the physical environment by simply being exposed to it. This is similar to what Erdelez (1997; see also 2005) called *information encountering* or “a memorable experience of unexpected discovery of useful or interesting information” (1997: 179).

Savolainen (2016) elaborated upon Bates’ classification, expanding upon the terms I have described above in order to make them more expressive and more precise. Especially when it comes to the description of the passive undirected behavior, Bates’ choice (“being aware”) does not seem quite fitting, as it describes less a behavior than a state of mind. Savolainen, conceptualized active modes of information seeking as: (a) active seeking and searching; and (b) browsing and scanning; and passive modes as: (c) passive monitoring; and (d) incidental acquisition of information:

Table 2: Modes of information-related behavior according to Savolainen (2016)

VARIABLES	ACTIVE	PASSIVE
DIRECTED	(A) Active Seeking and Searching	(C) Passive Monitoring
UNDIRECTED	(B) Browsing and Scanning	(D) Incidental acquisition of information

Both Bates and Savolainen worked on establishing a general conceptual framework for describing information-related types of behavior. I am not aware that this type of conceptualization has been applied before to the study of dictionary affordances

and dictionary-related behaviors. Yet I would claim that a consideration of this type of framework in the context of dictionary-related behaviors is a much needed corrective to the Function Theory of Lexicography.

Active modes of information-related behaviors from Bates and Savolainen are directly applicable to dictionaries: searching is aimed at satisfying recognized directed needs, whereas browsing is aimed at satisfying recognized undirected needs. When a user browses a dictionary, he or she is taking active steps at seeking information, but this behavior is not aimed at reaching one particular dictionary segment. Browsing is a type of exploration: a journey which is as important if not more important than the final destination.

Directed and undirected passive modes of information-related behaviors do not translate as easily to the dictionary context. There are three reasons for this:

1. In Bates' and Savolainen's general conceptual frameworks, the distinction between directed and undirected passive modes accounts for the possibility of incidentally acquiring information in two very different situations: (a) while being engaged in an information-seeking behavior (e.g. reading a newspaper and discovering an article on one's favorite writer even though seeking information about him or her was not the purpose of reading the newspaper in the first place); and (b) while being engaged in an activity whose main purpose is *not* to acquire information (speaking to friends, watching a movie etc.).
2. The distinction between directed and undirected passive modes is not always easy to make even within the general framework.
3. The scope of behaviors afforded by the dictionary *qua* lexicographic tools is much narrower than a general conceptual framework. My focus here is on

one information tool (dictionary), its realization in two different media (print and digital), and the kinds of behaviors that it makes possible for the users who engage with this tool. Surely, lexical knowledge is acquired in many situations not involving the use of a dictionary, but those lie beyond the scope of this investigation. That's why the distinction between directed and undirected passive modes is of little consequence for conceptualizing dictionary affordances.

Consequently, I am proposing the following conceptualization of information-related behaviors afforded by dictionary use:

Table 3: Information-related behaviors with dictionaries

VARIABLES	ACTIVE	PASSIVE
DIRECTED	(A) Search 1. primary 2. secondary	(C) Serendipitous discovery 1. Proximal 2. Relational 3. Incidental
UNDIRECTED	(B) Browsing	

I have subcategorized search into primary and secondary, whereby *primary search* refers to the directed, active behavior aimed at satisfying a user's information need which has been recognized as such and formulated *before* using the dictionary; and *secondary search*, which refers to the directed, active behavior aimed at satisfying a user's information need which emerges as a consequence of the primary search. The secondary search occurs, for instance, when a user follows a cross-reference from the entry that has been retrieved during the primary search; or if a user identifies an unknown word in the definition of the entry from the primary search, and decides to look it up as well. *Browsing* a dictionary is an active, undirected mode of information seeking which satisfies a user's general need for information,

accompanied by a sense of curiosity and, sometimes, pleasure (see Considine 1998), but without a formulation of a particular query, as is the case in primary and secondary search. Finally, *serendipitous discovery* is a passive information-related behavior in which the user comes across a notable piece of information without actively looking for it. Serendipitous discovery can take place *by proximity* (when a user notices a semantically related or unrelated entry close to the entry which he or she is reading); *by relation* when a given entry points the user via cross-references to another entry; or *by incidence*, when a user discovers a notable piece of information randomly, when opening a dictionary for the first time in a consultation session, or during browsing. A serendipitous discovery by relation can but does not necessarily need to be followed by a secondary search.

Regardless of whether serendipitous, accidental discovery of information takes place when the user is not looking for anything in particular or looking for information on something else, it is receiving increasing attention from researchers in information science (see for instance Lawley and Tompkins 2008; McCay-Peet and Toms 2010; Makri and Blandford 2012a; Makri and Blandford 2012b; Edward Foster and Ellis 2014; Agarwal 2015).

Serendipitous information retrieval occurs when a user acquires useful information “by chance”, i.e. without explicitly seeking it (Toms 1998). This type of information retrieval is to a certain degree influenced by the person’s prior knowledge but also by the recognition of *affordances* in the text itself (Toms 2000). Early experiments with dynamic links of suggested articles in an online environment have shown that users recognized “the need for and value of these chance encounters,” commenting in post-session experiments that “if you focus on your interests, then your interests are going to stay what they are.”

In the next section, I will describe more concretely how the general principles of information-related behaviors I have described are *afforded* by the dictionary, and what are some of the differences in the actionable properties of analogue and digital dictionaries.

1.3.4 *Doing things with and in dictionaries*

When discussing the affordances of a dictionary, I will limit myself to the situations of use related to acquiring linguistic, metalinguistic and general knowledge. I will not be addressing the non-essential uses of dictionaries, for instance as props for lamps, or as hiding places for love letters.²⁹ The recognition of the communicative or the metalinguistic need and the formulation of the query (what it is that the user wants to find in a dictionary) is also *not* the topic of this section. Neither are the material affordances of paper that are in principle not directly related to lexicographic access, such as leaving traces of engagement on a surface with writing or highlighting devices. The sole purpose of this section is to analyze the affordances of the dictionary as a specifically *lexicographic* tool, a tool that lets us engage with lexicographic data. For this, we need to presuppose that there is a user, with or without a specific information need, who intends to interact with a dictionary for information-related purposes. The main question here: what does the dictionary let this user *do* with it? How does it determine the possibilities of goal-oriented action?

A dictionary typically affords two basic types of action: *selecting* textual segments (entries) to be read and *reading* those same segments. The selection of a textual segment to be read can be achieved by *searching*, by *browsing* or

²⁹ This type of “improper” use is what Wiegand called “zweckentfremdete Benutzung” (unintended use; lit. use foreign to the purpose). See Wiegand 1998: 359; also 2000:741.

serendipitously. In a print dictionary, searching takes place by following the alphabetic order of the dictionary macrostructure (see Hausmann and Wiegand 1989), whereas in a digital dictionary, identifying the text segment to be read is usually achieved by typing a headword into a text field and retrieving the text segment from the database. For the sake of simplicity, I will be referring to these two types of activity as Searching (P) and Searching (D).

Searching (P) and Searching (D) have the same goal, but are based on different cognitive and motor skills (familiarity with the ordering of letters vs. using an input text field). Searching (P) is a more elaborate and more complex process of elimination, which involves:

1. *handling* a physical object;
2. *determining* where to start the search physically (where to open the book — towards the beginning, the middle or the end);
3. *decoding* running headers on each page, which indicate the first and the last lemma on the given page, to determine whether to move forwards or backwards to continue with the search;
4. *estimating* roughly the portion of the dictionary that should be skipped until checking the next set of running headers;
5. *deciding* when in the process the user is close enough to the entry so that it's no longer worth it to skip portions of the dictionary (to avoid going too far), in order to start
6. *turning individual pages* until reaching the page, which, based on the running header, must contain the entry the user is looking for;

7. *skimming* the bold-faced lemmas on the given page to identify the entry containing the lemma the user is interested in. Once and if successful, the user can start
8. *reading* the entry.

The entire process described above is heavily influenced by the materiality of the book (its dimensions, weight, the thickness of paper) and a complex interplay of abstract notions such as alphabetic order and the physical space that dictionary entries take up on any given page.

In lexicographic literature, it is commonplace to describe the process of Searching (D) as a liberation “from the straitjacket of ... alphabetical order” (Atkins 1996: 516). Searching (D) is, indeed, a more straightforward and less physically involved process for the user: it consists of:

1. *identifying* the search input field on a computer screen;
2. *focusing* on the search input field, usually by moving the computer cursor over it and clicking into it;
3. *typing* the query on the computer keyboard; and
4. *submitting* the query, usually by either clicking on an input button, or the return-key on the keyboard.

Once and if successful, the search result will be displayed on the page.

Searching (P) and Searching (D) are different both in terms of their cognitive and mechanical requirements. They also, however, afford different *reading* opportunities. The result of Searching (P) is not separable from the flow of dictionary content. Textual segments in a print dictionary remain fully embedded in the dictionary page: the user can effortlessly identify, read and compare the entries *before* and *after* the selected entry. This affords easier identification of both

semantically related neighbors (when they appear next or close to each other) as well as semantically unrelated neighbors whose proximity is due to alphabetical chance. In either case, a print dictionary easily and always affords *contextualized reading* and the user's engagement with the *uninterrupted flow of dictionary content*.

More often than not, the result of Searching (D) appears in isolation: the selected entry is displayed on the computer screen, but neither the previous nor the following entries are. Reading a dictionary entry digitally usually amounts to the reading of an extracted textual segment, decontextualized from its place in the dictionary hierarchy.³⁰ Some online dictionaries (such as OED, for instance) provide next to the entry proper a navigation bar displaying headwords for a number of entries preceding and following the displayed entry. In those cases, links are clickable and can take the user to the listed entries, but they, too, will be viewable in isolation from each other, even though they appear next to each other in the printed source dictionary. In other words, online dictionaries often afford no way of viewing, reading and comparing multiple entries at the same time. There are some exceptions to this, such as the online version of the *Grimm-Wörterbuch*. The Grimm interface provides a navigation bar *and* the content of several entries per page.³¹ The navigation bar, however, in the case of the Grimm Online, unlike the OED navigation bar mentioned above, shows only a number of headword links *following* the top entry in the given view, and no links to the previous entries. Previous entries can be reached by clicking on the button that takes the user to the previous page, but there

³⁰ See for instance <https://www.wordnik.com/words/pastoral> or <http://www.dictionary.com/browse/pastoral>.

³¹ See, for instance, http://woerterbuchnetz.de/cgi-bin/WBNetz/wbgui_py?sigle=DWB&mode=Vernetzung&lemid=GS24219#XGS24219.

is no indication on the navigation bar what those previous entries are. This is less convenient for the user than what is afforded by the printed edition where the concurrent viewing of the preceding and the following context are only limited by the physical space of two dictionary pages facing each other.

From the perspective of targeted information extraction, which is the only kind of extraction that Function Theory, for instance, deems legitimate and important, displaying entries in isolation poses no problem. If anything, it helps the user fulfill his or her specific information needs more efficiently and without distraction, but only during primary searches, i.e. in those cases where the user has very specific information goals. For different types of behavior — browsing and serendipitous discovery — many online dictionaries offer fewer possibilities for the user's engagement with the uninterrupted flow of dictionary content. This, as I will show in my discussion of the user interface for the digital edition of *LSGL* can have far-reaching consequences for the *undirected* acquisition of information from online dictionaries.

Both print and digital dictionaries afford identifying and following explicit links (cross-references) between different entries. In a print dictionary, users who want to follow a cross-reference to an entry which is not laid out on the same page as the source entry need to complete the Searching (P) process a to a target entry. In doing so, they have to either abandon the source entry altogether or they can physically bookmark the page of the source entry as they look up the target entry. In case the source and the target entries need to be compared, the users have to switch back and forth between two different physical locations in a dictionary, which in case of multivolume dictionaries involves placing several books next to each other, or stacking them on top of each other. A comparison of more than two

entries becomes a juggling act. In online dictionaries, on the other hand, cross-references are usually displayed as clickable hyperlinks, which makes the process of following cross-references much easier. At the same time, most online dictionaries will not allow concurrent display of two or more entries from the same dictionary, making the following of cross-references relatively easy, but comparison of different entries less so.

Thinking of dictionaries as tools which have their own affordances, i.e. allow certain types of user behaviors and actions, gives us a fresh view of both print and digital dictionaries. On the one hand, our discussion above shows that print dictionaries usually fare better than their digital counterparts when it comes to viewing entries in context because the layout of print pages affords engagement with an uninterrupted flow of content. Print dictionaries are easily browsable and fairly flexible when it comes to serendipitous discovery. Especially when it comes to affording serendipitous discovery by proximity, print dictionaries are still not matched by most online dictionaries. On the other hand, even though hyperlinked cross-references in digital environments provide quicker access to isolated entries, they usually do not create an added value in terms of affording easy comparison of different entries. Online dictionary interfaces are still in their prolonged infancy with very little new ground being broken. I will take up this issue again in Chapter 6 (pp. 217-250), where I will show how the previous discussion of dictionary affordances and information-related behaviors has influenced the creation of the API and the user interface for the digital edition of Vuk Stefanović Karadžić's *Lexicon Serbico-germanico-latinum* which has a strong emphasis on not only primary and secondary searching, but also on browsing and serendipitous discovery.

AN OVERVIEW OF THE HISTORY OF ELECTRONIC LEXICOGRAPHY

The problem with hindsight, historian Robert J. Young wrote in his analysis of France and the origins of World War II, is that “it is illuminated more by the present than the past” (Young 1996: 108). The present — in the sense of our collective experience — is not only our only possible vantage point but also, unavoidably, a distortion field through which we explore the past. When looking back at the history of electronic lexicography, it is all too easy to slip into techno-imperialist discourse: to assume automatically and uncritically that we are at the apogee of technological growth and that what came before us is of little consequence for what we do today. From the perspective of our networked, connected world, we can, indeed, smile at the naive thought of a lexicographer who in 1973 was “intrigued by the possibilities of the self-mailing postcard” (Chapman 1973: 309) as a technological solution for collecting lexical citations from volunteers; or his colleague who declared that the use of computers in lexicography “is at or near the point of saturation” (Venezky 1973: 287). But our sense of technological superiority and our hindsight bias will not help us get closer to an understanding of either the historical

moment in which such pronouncements were made or the true scope of previous technological achievements. What's more, our own sense of self-satisfaction may cloud our view and make us overlook some ideas from the past that could only now be brought to fruition.

The production of dictionaries has been for several decades hardly imaginable without the use of computer technology (Knowles 1989; Meijs 1992; Hockey 2000). Easily searchable large corpora, the ability to study collocational features, and the advent of the Internet as a platform for the production and distribution of knowledge have undoubtedly altered lexicographic practice. At the same time, the use of computers has not changed one fundamental aspect of lexicographic enterprise: a perceptual disjuncture between what Seargeant (2011) calls "lexicographic ideology" — a widely held belief that a dictionary represents an accurate and complete account of a fixed entity called language — and the reality of lexicographic practice which can in fact result exclusively in partial and incomplete snapshots of language, which, as an institutional and a social construct, is always a moving target. With or without digital technology, an ideal dictionary is merely a chimaera: it can only ever be imagined. An ideal dictionary would be impossible to produce — not only because language changes and new words are constantly being coined — but also because lexicographic knowledge, despite the appearance of authority, is, ultimately, based on interpretation.

Looking back at how other scholars employed technology and how they imagined ideal dictionaries of the future, both explicitly by projecting their visions of what electronic dictionaries should be capable of, and implicitly, by developing lexicographic formats, models and platforms, can provide important insights into the actual development of lexicographic and metalexicographic thought. Even more

importantly, perhaps, a retrospective look can have a truly sobering effect. It reveals the extent to which we are still, in a manner of speaking, in the early days of electronic lexicography. As we already saw in Chapter 1, dictionaries have been around for many centuries, during which certain methodological and ideological practices have led to the solidification of the tool and the concept we know as “the dictionary.” But “computerized” dictionaries have been around for some sixty years only, and globally available electronic dictionaries, dictionaries for the masses, so to speak, only since the advent of the World Wide Web. It would be wrong and short-sighted to believe (as some lexicographers did in the seventies) that we have already reached the pinnacle of technological development in lexicography. To look back is one of the ways we learn how to look forward.

2.1 DH and eLexicography: common roots

The use of computers in manipulating humanistic texts dates back to the concordancing of the works of St. Thomas Aquinas by Jesuit priest Roberto Busa (Busa 1980). While conducting research for his 1949 doctoral dissertation on the metaphysics of presence in St. Thomas Aquinas, Busa realized that a “philological and *lexicographic* inquiry into the verbals system of an author has to precede and prepare for a doctrinal interpretation of his works” (83, my emphasis). To create an *Index Thomisticus*, a concordance of all the words of Thomas Aquinas, he also concluded that he needed to “look for some type of machinery” in order to process more than ten million words. Busa managed to get IBM’s founder Thomas Watson on board, which was no small feat, considering that the punch card machines, the most advanced computer technology of the time, were not meant to process texts

but rather crunch numbers.³² The cumbersome nature of text input and the limitation in the number of characters that could be represented on each card would plague humanities scholars *and* lexicographers, for decades.³³ Yet the use of data-processing techniques which had been developed primarily for science and commerce would not only prove to be a significant factor in facilitating information retrieval and textual analysis, but would, as an IBM engineer and one of Busa's collaborators noted at the time, "initiate a new era of *language engineering*" (Tasman 1957: 256)

The first, provisional, results of Busa's cooperation with IBM came in 1951 when Busa presented his *First Example of a Word Index Automatically Compiled and Printed by IBM Punched Card Machines*. Over the next three decades, Busa and his assistants lemmatized and indexed the entire works of St. Thomas Aquinas, moving from punch cards to magnetic tapes and eventually producing, in 1980, the 56-volume *Index Thomisticus* on paper. The data was eventually transferred onto CD-ROM (1992) and a website (2005).

The chronology of the *Index Thomisticus* reflects the way changes in technology also affected early developments in electronic lexicography. What was

³² The first large-scale application of punched cards to process data was during the 1890 US Census thanks to a system developed by Herman Hollerith. After the success of the census, Hollerith continued to work on the system in order to get private business and public organizations to adopt new calculating methods. His Computing-Tabulating-Recording Company (CTR) was later renamed International Business Machines, or IBM. On the early history of tabulators and punch cards, see Truesdell 1965; Aspray *et al.* 1990; Norberg 1990; Heide 2009.

³³ The first cards Busa used had only 38 characters, but the number of characters in the standard IBM card in the sixties went up to 64. Needless to say, it was linguists and lexicographers who quickly discovered that inputting medieval and dialectological texts with diacritics required complex coding schemes: "for example, ã will be rendered by,A and ā by -A and ÄY by *0A*Y (0 being zero)" (De Tollenaere 1972: 148). For conventions used in recording and processing dialectological materials, see Kamp (1969) and Keseling *et al.* (1970).

originally a computer-assisted project to produce a paper-based reference work, first using IBM's punch cards and, later, magnetic tapes, eventually became an electronic product, easily available to anybody with Internet access. As the subsequent sections will show, a similar process, in which the computer medium turned from a tool assisting in the production of dictionaries to a tool for distributing and consulting dictionaries, has characterized the development of electronic lexicography as well.

2.2 The sixties: a brave, cumbersome new world

One of the earliest examples of the computer-assisted preparation of a paper dictionary was the *Random House Dictionary of the English Language*, whose production system was designed by Laurence Urdang (1966). The decision to make a large unabridged dictionary of about 260,000 entries (twice as many as the American College Dictionary of 1947) was made in 1959. Urdang and his team "hoped to use computers to do the sorting, codifying, rearranging, and checking the data at hand and the text to be written" (31). The challenges of processing lexicographic data in the 1960s were, on one level, quite different from our own: "one of the early problems in any computer program is getting the information into the computer" (31), but, at the same time, it was the nature of lexicographic data itself that posed difficulties of their own:

dictionaries create special problems. They are, both in information content and typography, probably the most complex books in the world. (31)

Using forms specially designed for this purpose, the 130-odd thousand entries from the ACD were punched into standard 80-column cards. Dictionary data was divided

into seven categories: (1) illustrations, (2) main entry words (including pronunciations, inflected forms), (3) definitions, (4) variations, (5) etymologies, (6) run-on entries (undefined words formed by the addition of suffixes to the main entry word, such as *-ly* or *-ness*), and (7) additional information like synonym studies, usage notes etc. (Ibid.: 31-2). The definitions were furthermore annotated using 158 subjects fields (such as *PHYSICS*, *CHEMISTRY* etc.). The explicit terminological labeling made it possible to retrieve each of the subject fields for specialists to review the copy. The ACD data was converted to paper tape and printed out for editors to work with. More than 200 outside consultants reviewed the contents of the new dictionary, and, after another set of corrections, the contents were sent to operators who punched the contents of the RHD dictionary back onto 700,000 feet of paper tape (32). Separate computer programming firms were employed to convert the paper tape to magnetic tape and produce the final alphabetic order of the dictionary.

Originally, the editors planned to have the dictionary typeset automatically, but “there were no firms large enough or versatile enough or confident enough of their equipment to convince our manufacturing department that such an undertaking would be feasible” (33). That’s why the entire manuscript was converted to more than 21,000 frames of microfilm, which were then enlarged and photocopied on 12” by 12” paper, and individual sheets bound into some 70 books. After a final two rounds of copy-editing, the material was sent off to the compositor for typesetting.

The effort that went into planning and executing a project of this magnitude was enormous. Data input was difficult and time-consuming, but Urdang claims that the sorting and merging capabilities of the computer saved 150-man

years of labour to the team (33). Even more interestingly, however, Urdang mentions that the very nature of the processing “demanded a codification of thinking on the part of all who were associated with the RHD.” Dictionaries have always been highly structured texts, but the use of computer technology in the processing of lexicographic data underlined the need for strict consistency in the way the lexical material is assembled, organized and presented.

It was the structural underpinnings of lexicographic works and a wealth of linguistic data contained in them that led scholars to explore the idea of producing the first machine-readable versions of already published print dictionaries as well. With the initial goal of developing a formal semantic description of the English language consisting of a set of lexical entries and a system of rules to determine appropriate readings of both individual sentences and connected texts, Olney *et al.* (1968) transcribed two English dictionaries, Webster’s Seventh New Collegiate Dictionary (W7) and The New Merriam-Webster Pocket Dictionary (MPD), into machine-readable forms and developed applications to use them in computational lexical research (see also Revard 1968). A study of word frequencies in dictionary definitions, for instance, would empirically confirm that the most frequent words used to define senses in these dictionaries (terms such as SUBSTANCE, CODE, THING, KIND, POSSESS etc.) were practically identical to the semantic primitives in AI or in linguistic theories, such as those of Fodor and Katz (Wilks *et al.* 1996: 82). Areas of study in which the first MRDs would eventually be applied included taxonomy extraction, text analysis, speech processing, syntactic and semantic parsing, detection of circular definitions etc. (see Amsler 1984)

After W7 was converted to a machine-readable format in 1968, Donald Sherman developed an elaborate lexicographic data format to standardize the

representation of dictionary entries (Sherman 1974). He based his format on MARC (Machine Readable Catalog) — the Library of Congress standard for exchange of bibliographic data (Avram *et al.* 1965) and called it WEBMARC. Sherman's WEBMARC used the MARC-like scheme to tag individual components of a dictionary entry (such as spelling variants, etymologies, definitions etc.) and make them accessible independent of each other. The subfield structure of pronunciations was detailed enough, for instance, to label individual consonant and vowel segments. With this type of mark-up, it was possible to use the computer to search for patterns of phonetic data and answer questions such as: what English words exhibit vowel tensing in a syllable preceding a bivocalic sequence (e.g. area, gymnasium etc.) (23)

The promise of dictionaries stored as computer files did not only have to do with general access to individual dictionaries — or the ability to search and filter their contents for particular queries. The true promise of machine-readable dictionaries was seen in their extensibility and transferability:

One of the advantages of computer files over printed texts is their capacity to absorb additional information as a result of being compared and merged with other data files. (Sherman 1974: 25)

With the lexicographic material in computer-readable form, the history of a series of dictionaries may be traced in what may be described as a vertical dimension (by comparing several records of the same text) or in a horizontal dimension (by comparing the interconnections of several different but related texts). (Huntsman 1978: 55)

Sherman hoped that his W7 could be expanded with information from other dictionaries, such as the machine-readable *Chronological English Dictionary*, based on the earliest usage information recorded in the *Shorter Oxford English Dictionary* (Finkenstaedt *et al.* 1970) or the augmented tape version of the Pronouncing

Dictionary of American English (Kenyon and Knott 1953), which was developed at the Speech Communications Research Laboratory in Santa Barbara.³⁴ Sherman clearly understood the potential of digitizing other dictionaries: he explicitly mentions Daniel Jones' *English Pronouncing Dictionary*, which, if converted to WEBMARC, would provide a basis for a systematic lexical comparison of British and American pronunciations. But he also stresses the need for digitizing older dictionaries of the English language because of their "great linguistic value" and because "historical data can best be analyzed and compared in computer files" (25).

The 1960s computer methodology was employed in numerous scholarly dictionary projects, including the field of historical lexicography (Cameron *et al.* 1970; Aitken 1971; Leyerle 1971; Huntsman 1978). English-language lexicographers were not the only ones to realize the potential of digitizing textual material and using computer technology in the production of dictionaries. The *Trésor de la Langue Française*, a historical dictionary dealing with the period from 1789 to 1960 was based on a computerized file of 90 million words at the University of Nancy (Bernard *et al.* 2004); the Lexical Archive of the Italian Language was created at the Accademia della Crusca in Florence and used in the production of the *Dizionario storico integrale della lingua italiana* (Duro 1966; 1968; 1973; Zampolli 1973); while the *Dictionary of the Older Scottish Tongue* (DOST), developed at Edinburgh, covered the use of Scots words from the twelfth to the end of the seventeenth centuries based

³⁴ The computerized Chronological English Dictionary, which re-arranged some 80,000 headwords from the Short Oxford English Dictionary in order of their earliest recorded usage was later used in a diachronic statistical study of vocabulary development in English (Wermser 1976) and in the study of sixteenth-century English words present in Shakespeare corpus (Neuhaus and Spevack 1975).

on the Older Scottish Textual Archive of computer-readable texts (Aitken and Bratley 1967; Bratley and Lusignan 1976).³⁵

Computer specialists in the 1960s were very much aware that the use of computer technology in the preparation of dictionaries was still cumbersome. The original system — called INFOL — designed for storing, retrieving, and editing data for the *Dictionary of American Regional English* (DARE), therefore, assembled together various computer-aided components to allow retrieval of collected data from a central file and “on-line” editing “with a display scope and keyboard” (Venezky 1968: 72). The concept of a dictionary which would be stored in the computer, edited, analyzed and consulted via a terminal, which Venezky initiated in 1965, was to produce a publication-ready file to be sent to the printer, saving overall costs and “allowing a degree of cross-indexing and categorization that has never been achieved in lexicographical publications” (79). After a few years of trial and error, it became apparent, however, that Venezky’s sophisticated program was ahead of its time: it could neither handle the large amounts of data nor truly work on the hardware which was available at the time (Henderson 1974). In 1969, the process of conversion to a simpler program had started; and Venezky’s idea of machine-scanning typed sheets was abandoned in favor of punch cards (Ibid.: 120).

³⁵ It is beyond the scope of this brief snapshot to go into more detail of the early days of computational work on dictionaries, even though a lot of work that happened in the sixties and later provides a great deal of important materials for a yet-to-be written cultural history of electronic lexicography. See for instance, work on the preparation of a Hebrew historical dictionary (Ben-Hayyim 1966); early Russian work on “automatization” and “statistics” in lexicology and lexicography see Фрумкина 1964; Засорина 1966; Москович 1966; a plan on the “mechanical” processing of an etymological dictionary of Hungarian, see Папп 1968; a proposed “computerized” dictionary of Andean languages (Wölck 1969); a report from a roundtable on historical dictionaries held in Florence on May 3-5, 1971 (de Tollenaere 1971) etc.

Regardless of the initial setbacks in the use of advanced computer technology, DARE was an ambitious lexicographic project that could realistically be accomplished only with the use of computers. The dictionary staff collected data on word usage from 2752 informants, along with information about the informants themselves. The canonical form for a stored record (called “element”) consisted of 13 separate items: a unique numeric identifier, headword, pronunciation, grammatical class, variant spelling, geographic location, informant type (education, occupation and age group), source,³⁶ sense type (or semantic classifiers)³⁷, definitions, citations, usage data and notes. Multiple-word expressions were entered under each significant word in the phrase (e.g. *snake in the grass* was entered under both *snake* and *grass*, but the full treatment was provided in the first record and cross-referenced in the note elements from the second). The whole system was flexible enough to allow retrieval and sorting based on each of the fixed fields so that complex searches (called “interrogations” at the time) were possible: finding, for instance, all elements for which the informant had no more than a high school diploma and was classified as old.³⁸

In terms of establishing and recording connections between nonidentical items in the central file, the INFOL explicitly encoded (using Venezky’s terminology

³⁶ The sources were: Wisconsin English Language Survey (a study of Wisconsin speech from 1950), DARE questionnaire, DARE tape recordings, the reading program and the Linguistic Atlas (a study of American speech which had started around 1930). (Venezky 1968: 72)

³⁷ In case of *snake in the grass*, the sense types are *DECEIT* and *ADAGE*.

³⁸ “Within the pale of traditional lexicography, an editor is limited by a fixed filing system, within which record classifications, other than by alphabetic sequence and sometimes by subject, are forever entombed. Even in some of the recently publicized computer-aided schemes, only a small number of fixed fields can be designated for sorting - all other relationships being lost.” (Ibid.: 75)

here): (1) “definitional identity” — a relation of synonymy: “*Tote, pail* and *bucket...* are different names for the same physical object” (Ibid.: 78); (2) a “sense class identity” — or, in other words, semantic classification of individual senses such as “domestic animals or farm implements or drag-racing procedures” (78); (3) “etymological identity” — words with common origin, such as, for instance, “spelling pronunciations or French loan words”; and, finally, (4) “functional identity” — words sharing a sociolinguistic, stylistic or pragmatic function such as euphemisms or exaggerations).³⁹

2.3 The seventies: imagining the future

2.3.1 *Technology as a double-edged sword*

In 1973, the New York Academy of Sciences hosted a lexicography conference which brought together some of the leading figures of the field and included a section dedicated to the use of technology in lexicography (McDavid Jr. 1973). The conference was an opportunity for lexicographers to take stock of ongoing efforts in the field, but also to look toward the future. By this time, there is practically no doubt among practitioners in the field that the future of dictionaries will be electronic: despite the limitations of the technology at their disposal and the difficulties involved with text input and output, lexicographers agree that the dictionaries of the future will greatly benefit from the availability, searchability and manipulability of electronic texts. At the same time, however, the increased mechanization of lexicographic production is seen as a double-edged sword. The

³⁹ “Except for sense class identity, which could be pursued into complete chaos, all of these connections could be discovered if enough editors worked enough years with the data (or if enough monkeys poked randomly at classification keys for a sufficient infinity).” (Ibid.)

standardization and formalization of lexicographic data is believed to be an asset because it can lead to, for instance, better indexing and, therefore, more complex and mutually compatible resources, but all this at the cost of glossing over various irregularities and inconsistencies of language. The plans for the creation of a “Central Archive for Lexicography in English” reveal, however, at quite an early stage in the development of electronic lexicography, a strong need for the consolidation and integration of lexicographic resources in view of what Bailey would refer to as “socialized lexicography” (Bailey 1973: 296).

So what were the features of an “ideal dictionary” from the perspective of researchers in the seventies? In a playfully titled contribution “How to Make a N.U.D.E (New Utopian Dictionary of the English Language),” Revard (1973) engages in a stimulating exercise of lexicographic wishful thinking: he assumes, for the sake of the argument, a “Utopia in which lexicographers will be given unlimited funds, perfect facilities, innumerable expert staff, and sufficient time, and asked to produce the best dictionary they can imagine” (91). He focuses on two main requirements for such a phantasmic dictionary: (1) the need for lexicographic data to be in computer-accessible form so that it can be studied “backwards, forwards and inside-out,” and (2) the need to offer preliminary analysis of the possible types of semantic relations between different word senses. Machine-readable data should not only include citation slips but also “*the previous dictionaries in the language*” which should be mutually aligned into “full concordances” (91). This, according to Revard, will improve the consistency of lexicographic definitions, their style and format.⁴⁰

⁴⁰ Revard, who was at the time already an accomplished scholar and teacher of medieval English literature at Washington University in St. Louis, Missouri, specializing in Middle English, history of the English language and linguistics, would eventually become known for his poetry with Native American themes (Arnold 2007). In February 1967, he took a sabbatical from his

For Revard, the manipulability of the electronic text is a major asset: he envisages a scenario in which an electronic version of the “properly formatted” *Oxford English Dictionary* would be available to the researcher along with a “properly prehensile software” which could “turn the OED inside-out in a number of ways”: by ordering the data according to chronology⁴¹ or by usage labels (*slang, colloquial, figurative* et cetera). But Revard doesn’t stop there. In terms of the genealogy of lexicographic knowledge, his coupling of machine tractability and manipulability of individual bits of dictionary data, on the one hand, with a clearly articulated need to record semantic relations between different dictionary senses, strikes one as not only very forward-looking but also indicative of a deep engagement with the complexity of the dictionary genre. What Revard wants from an exemplary if fictitious dictionary is not just to be a properly indexed relational database in the modern sense of the word: he wants it to be a “lexical network” (92).

Revard admits, however, with self-irony, that in his Utopia there are no copyright problems — the issue which is also echoed in Lehmann (1973: 312). There is great optimism that the consolidation of existing information will be not only

teaching duties to work on a “computers and dictionaries” research project in Los Angeles, funded by the US Department of Defense, together with John Olney, among others. Years later, in an interview, he would describe his position as that of a “semanticist linguist” and the project as exciting and groundbreaking: “The job was to try and do something that I had been developing here—ideas about putting a large dictionary of the English language into computer accessible form, developing programs that would access it, things that are commonplace now, but this was pioneer. This was cutting edge. This was the state-of-the-art. This was the very biggest, strongest, and finest computers and some of the best linguists of the time. So I was very excited to be asked to do it. I was flattered and honored and over my head.” (Revard 2001a: 4). Revard’s paper at the 1973 conference in New York was his second that came out of this project. The first one he delivered in 1968 at a conference of the Association for Computational Machinery in Las Vegas (Revard 2001b).

⁴¹ Possible query scenarios include “all the word-senses that had disappeared by a given date, say; or all the senses that appeared between any two dates; or all the words that had as many as ten different senses between 1600 and 1625.” (Revard 1973: 92)

possible and desirable, but almost inevitable: “Proprietary rights notwithstanding, it is clear that commercial dictionary houses will shortly be obliged to enter into some such central arrangement” (Bailey 1973: 295).

Revard and his colleagues did not see the use of computers in the production of future dictionaries exclusively as a time or cost-saving measure. Speed was certainly considered as one important aspect of what computer technology had to bring to the table, but equally important was the realization that technology imposes structure on data and conditions research output: “the outcome of our research will be conditioned by the devices we use in its execution” (Ibid.: 293; echoing Richards 1955). Bailey points out, however, that this is not something unique to computer technology as such: any technology employed in the processing of lexicographic data is bound to leave its traces in the end product — whether one is talking about the use of databases or the pigeon-holes purpose-built to hold lexicographic slips in the preparation of the OED (see Murray 2001: 136).

Most scholars at the time would have agreed with Lehmann that machine-readable lexicographic data “will be will be increasingly formalized” in the future (Lehmann 1973: 312). Words are not numbers and dictionary entries are not mathematical equations. Computers need explicit instructions on what to do with particular sets of strings. In fact, computers need to be told what a particular set of strings is, in the first place: what specific characters constitute the lemma, what strings represent the part of speech etc. The explicit encoding of textual information and the explicit structuring of dictionary entries makes it possible for the computer to manipulate that data and to provide advanced search possibilities.

According to Revard, the formal aspects of dictionary definitions and a general attempt to create structured data can actually be quite beneficial in the

study of language precisely because not everything that occurs in language can easily fit regular structures:

the effort to provide completely regular and reasonably formal definitions in which usage at any point is consistent with usage at any other, is bound to bring to light anomalies, irregularities, inherent in the language itself, and these are likely to be precisely the sort useful to linguists who want to find the holes as well as the nodes of the lexical network. (Revard 1973: 91-2)

According to Revard, the increased level of formalization can lead lexicographers to discover exceptions which are interesting in their own right. In other words, systemic thinking about language is interesting both because of the parts that fit the system (the “nodes of the lexical network”) and those that clearly do not (“the holes”). A failure of a system is part of its own success.

At the same time, the depth of formalization is recognized as an investment into the productivity of the dictionary, so much so that Richard Bailey, discussing the potential of a proposed project for a Dictionary of Early Modern English Pronunciation, which would be based on the information stemming from the works of orthoepists and phoneticians of the time, criticizes the authors for not being ambitious enough and for treating the computer storage as a mere equivalent of traditional slips. He stresses the need for “sufficient forethought” and “indexing every element of their transcription rather than just the citation form that happens to occur” thus anticipating the possibility of complex question-answering systems.⁴²

⁴² The types of question that Bailey wants the system to provide answers for are: At what point does the evidence begin to show the shift from /e/ to /i/ in such words as tea? or Does preconsonantal and word-final /r/ begin to lose constriction after central vowels sooner than before peripheral ones?

the new technology allows us not only to disseminate lexicological information... but also makes it possible for us to anticipate richer results from our collection of data than have emerged from less adaptable and convenient schemes using the old technology. (Bailey 1973: 296)

At the same time, Bailey raises an enormously important issue about the nature of lexicographic structures: the contrast between what he calls “linguistic facts” and “life of the community:”

Linguistic facts, in our historical dictionaries, are left somehow apart from the life of the community; we have yet to capture in our representations of meaning the continuum from precision to imprecision inherent in our citations and certainly pervasive in the collective whole. (Ibid.: 294)

Bailey quotes the attempt of Charles C. Fries in a 1932 sample of his *Early Modern English Dictionary* to introduce a category for “imprecise” citations where it is “impossible to determine the precise meaning of the word.” Fries subsequently abandoned this idea after being criticized for it:

the technology of pigeonholes prevailed and the tradition continues that, whatever the imprecision of some speakers, the English language that our dictionaries represent is really a pure light illuminating the darkness: “Life, like a dome of many-coloured glass, / Stains the white radiance of Eternity. (Ibid.)

Bailey is concerned that the use of technology can actually damage the integrity of data. When lexicographers quote only those sources where the word in question is easily attributable to a particular word sense, they are making their data fit a system which is much neater and better organized than language itself.

Bailey’s argument, however, does not only reflect his misgivings about sense disambiguation as such. His argument touches upon an even more fundamental

question about the role of human intervention in the automatic processing of lexical data:

[We cannot] reasonably expect the computer to execute the direction that Murray issued to the volunteer readers for his dictionary: ‘Make a quotation for every word that strikes you [Oh machina] as rare, obsolete, old-fashioned, new, peculiar, or used in a peculiar way.’” (Ibid.).

Dictionaries are reference works. They may have initially, as we saw in Chapter 1, provided only information about the so-called “difficult words,” but over time dictionaries assumed the role of lexical know-it-alls: providing as complete a picture about the vocabulary of a given language as is humanly possible. In earlier epochs, the dictionary as a textual genre was also fostering a sense of estrangement, emphasizing not so much the typical, as the unique, the exceptional, the unusual. The manual selection of quotations from reputable sources, such as those used by Johnson, for instance, made the dictionary not only informative but also highly enjoyable — if, admittedly, somewhat peculiar — literary genre.

It should come as no surprise that the lexicographers in the early seventies, working in a field not yet dominated by corpus lexicography, frequency lists and statistical passions, are still very much confident in the primacy of linguistic intuition:

I’m sure that these machines, no matter what champion speed-readers they may be, will not ever be able to supplant the human reader with his sensitivity and Sprachgefühl. (Chapman 1973: 309)

Scholars emphasized the continued significance of the lexicographer in what was already becoming a computer-dominated discipline because they felt the computer

could replace neither the “selfless volunteers of bygone days” nor “Murray’s pigeonholes so delicately imposed on the lexicon” (Bailey 1973: 294).

It is this same kind of championing the humanistic aspects of dictionary production which lead Venezky to conclude that:

the evolution of computers into the mundane tasks of lexicography is at or near the saturation point, that is, we’ve gone about as far as we can go in developing new applications under the current rules of the game, and any further qualitative advantages to be offered by automation will require a revolution in both semantic analysis techniques and in the interaction between lexicographer and computers. (Venezky 1973: 287)

Venezky sees no reason to expect that faster and more sophisticated machines will necessarily produce better dictionaries because available technologies were already producing more data than an editorial staff could actually process:

the present use of computers in lexicography offers no qualitative advantages - either now or in the future - to the users of dictionaries. (Ibid.: 290)

The change is evolutionary, but if there was ever to occur a revolutionary shift, it would, according to Venezky, not involve data transfer (from print to a computer readable format) and data matching (sorting and alphabetizing) but rather computer-assisted analysis (e.g. lemmatization, separation of homographs, sense classification) and “integrative processes” (composing definitions, selection of illustrative contexts etc.) (Ibid.: 288).

2.3.2 *The missing infrastructure*

Despite the fact that lexicographers in the seventies could not imagine the scope of technological change that would be ushered in by the advent of the internet, they were already articulating a need for infrastructural services that would provide

high-quality data to researchers. One such infrastructural service was Barnhart's proposal for a "Central Archive for Lexicography in English." It envisioned a collection of 25-30 million quotations for some 500,000 lexical items of written international English selected on the basis of frequency, range and "cruciality" — "the critical importance of the basic vocabulary of a particular field of knowledge and has little to do with frequency or range" (Barnhart 1973: 302). Barnhart relies on the "proper coding" of information so that relevant information can be extracted by the lexicographer or researcher:

The mere collection of data without purpose is meaningless; if we do this we merely follow the first law of computerdom: garbage in-garbage out (Ibid.).

It would be wrong to see the proposed archive as an equivalent or even a direct precursor to linguistic corpora that dominate the field of lexicography today. Corpora are searchable collections of authentic texts: they are not meant to be artificially created by linguists to fit a predefined theory; nor are they supposed to be tainted by the personal idiolect, intuition or introspection of the researcher (see Barlow 2011). Corpora constitute, to a large extent, raw data, although the process of selecting texts for a corpus and the very notion of a balanced corpus are far from mechanical and random. The proposed "lexicographic archive" could be best described as a corpus of annotated excerpts specifically geared towards illustrating individual lexical items:

The archive will have to assemble phonological, paradigmatic, syntactic and semantic information for lexical items, and list the syntactic and semantic properties of constituents required in the syntactic environments of these items. It will have to establish the implication relations between

such properties. It will also have to establish the logical relations between the meanings of linguistic items. (Lehmann 1973: 317)

This depository “where qualified collectors may send quotation files” (Barnhart 1973: 306) should be “systematically voracious, omnivorous in its appetite for linguistic material” and “an ingeniously computerized bank usable for research in grammar as well as in the more traditional concerns of lexicography” (Chapman 1973: 307).⁴³

Even though the proposed archive owes much more to the tradition of lexicographic slips than the emerging field of corpus linguistics, the lexicographers at the time were clearly aware that the benefit of access to large bodies of text and extended quotations⁴⁴ would create dictionaries capable of taking “more careful account of larger structures” and “freed in part from the uncomfortable tyranny of the isolated term” (Chapman 1973: 307).

Long before the Internet, before the participatory turn of Web 2.0 and social media, lexicographers realized the potential of computer technology for scholarly communication and collaboration. “Socialized lexicography, in short, is now upon us,” Bailey proclaimed, referring to capacity for “sustained and continuous growth through addition (or deletion) of information” and the chance of scholars from different institutions to work on various dictionary projects together (Bailey 1973: 296). Still important to this day is Bailey’s earlier call to data sharing:

⁴³ One decade later, commenting on the use of large databases for information retrieval such as LEXIS-NEXIS for lexicographic work, Landau (1984: 283) complained that such systems bury “significant facts beneath mountains of trivial variation, redundancy, and irrelevance.”

⁴⁴ Barnhart made “a plea for including enough of the context to give some idea of the field of operation of a word. Dictionary citations often merely prove the existence of a meaning; quotations show its use and are often meaningful in their own right.” (Barnhart 1973: 305).

For a relatively low cost, the citation file itself can be published on microfiche, making the evidence - if not the interpretation of the evidence - available to scholars before a generation or two has passed (Bailey 1969: 171-2)

The format du jour may have changed — microfiches have no place in the age of cloud computing — but the need for open access to scholarly data has remained an essential requirement for the creation of digital research infrastructures.

2.4 The eighties: toward standardization of data formats

2.4.1 An emerging discipline

With an increasing number of projects using and producing machine-readable dictionaries (MRDs) as repositories of grammatical, semantic and encyclopedic information, the eighties saw the emergence of the computational lexicography “as a discipline in its own right” (Walker 1989: 2). The fact that lexicographers need large bodies of diverse texts in order to adequately describe language was becoming more widely accepted and acted upon. At the same time, it became increasingly clear that a large number of incompatible data formats could be a major obstacle for the development of the field.

Textual corpora were not new to the eighties: some, like the Brown Corpus of American English, had already been around since the sixties (Kučera and Francis 1967), and its British English counterpart, the Lancaster-Oslo-Bergen Corpus, since the seventies (Johansson 1985). But the eighties saw the completion of a major milestone: the compilation and the publication of the *COBUILD Dictionary*, the first print dictionary that was in its entirety corpus-based (Sinclair 1987). The

interaction between texts and dictionaries, or *sources* and *resources* was recognized as essential (Walker 1986).

The lexicons for natural language processing (NLP) were at the time mostly illustrative and not large enough, with very little consensus on what the nature of lexical data is or how it should be represented (see, for instance, Ingria 1995). While quite large lexicons had been developed for specific application such as syntactic and morphological parsing (Sager 1981; Heidorn *et al.* 1982; Bates *et al.* 1986; Russell *et al.* 1986; Ritchie *et al.* 1987), the systems developed were hardly compatible with each other and would often end up using different representations for similar types of information (Boguraev and Briscoe 1989). This made the task of sharing data between systems difficult, if not altogether impossible. The existing MRDs, however, proved to be useful for various tasks such as exploring lexical relations and extracting taxonomic structures (Amsler 1980; 1981; Ahlswede 1985; Chodorow *et al.* 1985; Evens 1988; Ahlswede and Evens 1988).

One MRD in particular, the *Longman Dictionary of Contemporary English* (LDOCE) (Procter 1978) took up a prominent position as the favorite object of computation. The choice of LDOCE as *the* dictionary for computational analysis was propelled by two factors: the fact that the dictionary publisher agreed to offer its computer files to researchers; and the fact that the very structure of the LDOCE as a dictionary aimed at foreign learners of English made it a particularly suitable candidate for the automatized, computation extraction of data: (1) LDOCE employs an elaborate system of grammatical codes at both entry and sense levels that explicitly classify its lexemes beyond the traditional labeling of part of speech (noun, verb etc) by introducing explicit markings of valency (e.g. noun followed by THAT clause); (2) the computer files of LDOCE (unlike the printed version) also

contained a system of semantic codes that restricted subjects and objects of verbs, or nouns that adjectives can modify, specifying, for instance, that *persuade* requires a [+ HUMAN] object; (3) LDOCE's definitional vocabulary was restricted to some 2,000 items which facilitated the syntactic and semantic processing of the dictionary definitions (for more details, see Michiels *et al.* 1980; Boguraev and Briscoe 1987).

In addition to the use of dictionaries for developing NLP resources, the eighties also saw the first successful efforts in the creation of MRDs for direct human consumption. Boguraev *et al.* (1987) argued not only that “placing the dictionary on-line in an environment which supports fast interactive access to data selected on the basis of a number of linguistic constraints” (63) represents a qualitatively different mode of using MRDs than the then prevalent processing of dictionary data for later re-use in NLP applications, but also that “on-line” dictionaries have “the potential of making maximal use of the information typically found in a machine-readable dictionary”⁴⁵ (63). Data extraction for NLP purposes was based on preprocessing the publisher's source tapes for particular tasks: for instance, matching the LDOCE grammar codes into feature clusters to be used by syntactic parsers (Boguraev and Briscoe 1987). The “on-line” use required the segmenting and pre-processing of the source tapes to support more than one application:

Thus we should aim at incorporating the segmented version of the source intact into the database, to serve directly as its ‘bottom layer’ in the sense that all access paths ultimately point to complete dictionary

⁴⁵ The term on-line refers here to a local network of workstations rather than the global network that the Internet would become.

entries, which are then returned as the results of queries. (Boguraev *et al.* 1987: 65)

While Boguraev's LISP-based dictionary system took advantage of the interactive graphic functionalities of Xerox's LISP workstations, which made it possible for the user to construct complex queries using predefined dropdown menus, it stopped short of full browsing capability.⁴⁶

The eighties also saw the completion of the machine-readable version of the largest dictionary of the English language, the *Oxford English Dictionary*. Twelve large volumes (containing 41.81 million words in 252,259 entries) plus four *Supplements* (with additional 69,372 entries and 14.5 million words of text) were manually keyboarded "with insertion of typographical tags to capture the most salient characteristics of the entries." (Raymond and Tompa 1988: 872). Subsequently, a parser based on finite-state automata was built to complete and verify the tagging process, and to convert typographic to semanticized structural codes (Kazman 1986). In parallel, work was done on the creation of a database that would give users a chance to search and engage the dictionary as a hypertext (Tompa 1992). The decision to "computerize" was based on both economic and editorial grounds as the sheer volume of data made revisions and new editions of the dictionary financially unviable. At one point, OUP estimated that it was losing around one million pounds for each print volume of the *Supplements* (Gray 1986: 46).

⁴⁶ A typical example of a dictionary query that Boguraev cites is "the user wishes to see all entries for three-syllable nouns which describe movable solid objects, whose second syllable has a schwa as peak, and whose third syllable has a coda that is a voiced stop" (Ibid.: 67), which still indicates the use of the MRD for linguistic research more than general dictionary use.

Once again, as we saw (p. 84) with the example of the “self-mailing postcard” in the seventies, the current state of technology was shaping the expectations and considerations of success:

consider that the raw data in the *New OED* represents 500 million bytes (half a gigabyte) and that it can be easily contained on one of the new laser-read 4¼ inch compact discs (CD) used in the virtually flawless reproduction of music. The database designers estimate that the complete database could now be easily contained on three or four such discs. And as data compression increases, probably fewer discs will be required by the time the electronic version is ready. Soon microcomputers will be able to use the CD in special disk drives giving the PC a huge read-only memory. And who knows how long before the PC will be able to write on a CD? (ibid.: 49)

The “huge” memory of a compact disc and more than modest dreams of writable CDs cannot hide the fact that the prospect of “computerized” lexicography for end-users was already in sight.

2.4.2 *Common standards*

The difficulties of converting typesetting tapes to a format that could be computationally exploited or consulted by humans in user-friendly ways was largely a problem of parsing a flat character stream into a structure which explicitly represents various individual components of a dictionary entry (Alshawī *et al.* 1989: 43). The ambiguity of flat character sequences is relatively easy to decode if you are an experienced dictionary user, familiar with the conventions of the genre and a particular dictionary that you are holding in your hands. This is much more difficult for an algorithm for two reasons: (1) control characters that mark the beginning of a particular typographic scope (such as bold, italic etc.) do not necessarily have an unambiguous ending character counter-point: they could end with a switch to any

other scope (a different font, a superscript character etc.); and (2) conversion from typographic to structural tags is not unambiguous, because certain classes of typographic features can be used for multiple structural elements in a dictionary.

The development of common standards for the representation of lexical data was therefore seen as a *sine qua non* from the early days of manipulating lexical data with the help of computers:

if a standard data structure were to be used as a common defining medium for different record formats and data bases, then this common structure could serve as a unifying basis for exchanging and integrating related materials. A common data structure would at least standardize our vocabulary for describing and documenting data base format and content. (Sherman 1974: 21)

Before one could agree on a common standard for describing lexicographic data, more work had to be done on standardizing the representation of textual data in general. Goldfarb and his team at IBM proposed (1970) something they called Generalized Markup Language (GML), an abstract metalanguage capable of describing the structure of various kinds of documents without restricting them to a single formatting style or processing system. GML was based on the following two postulates:

1. Markup should describe a document's structure and other attributes, rather than specify processing to be performed on it, as descriptive markup need be done only once and will suffice for all future processing.
2. Markup should be rigorous, so the techniques available for processing rigorously-defined objects like programs and data bases can be used for processing documents as well. (Goldfarb 1981: 69)

GML was developed not as an academic exercise in debating textual hierarchies and ontologies, but for very practical reasons: as a way for ensuring consistent, efficient and rigorous processing of textual data. This is a point well worth remembering. When we speak of descriptive markup today, the separation of structure from presentation is always mentioned, and rightly so, as its most important defining feature. But structured documents do not exist in a metaphysical vacuum. They are — and have always been — part of various applications and processing systems. Documents are structured so that we can *do* things with them. This will be the guiding thought behind my encoding of the *Lexicon Serbico-Germanico-Latinum*.

GML used so-called generic identifiers (GI) delimited by a colon (:) at the start of a textual segment that was being marked up and a double colon (::) at its end. Furthermore, a colon would separate a GI from any text that followed. A paragraph would start with a GI that looked like **:p.** and end with **::p.** Furthermore, GIs could be further qualified with attributes containing processing instructions, for instance: **:artwork depth=24p** (Ibid.).

GML was the basis for the development of SGML (Standardized General Markup Language), which became the International Standard Organization's standard (ISO 8879) for document description in 1986 (see Goldfarb and Rubinsky 1990), which later led to the creation of a W3C standard markup language, the Extensible Markup Language (XML) which would eventually become the lingua franca of data exchange.

2.5 Conclusion

The use of computers in and for lexicography since the nineties is much better documented and better known than its early history. I will not be reviewing the developments since the nineties here, although I will be returning to the topic of markup and its importance for the scholarly editions of legacy dictionaries in Chapter 3 *Data modeling as an interpretative framework* (pp. 113-128). The goal of this chapter was not to trace a history of computational methods in lexicography, or to tell a self-congratulatory narrative of technological progress. Quite the contrary. The above discussion shows that a history of electronic lexicography — a history yet to be written — will have a lot to teach us: not only about how far we have come, but also how far we still have to go. We are yet to put into practice a truly “socialized lexicography” — an infrastructure that would make it possible for scholars to collaboratively update, annotate, interlink and discuss dictionaries. We are yet to find a sensible, research-friendly solution for problems related to copyright. And we are yet to make significant advances in the study of lexicographic complexity and ambiguity in the electronic age. Just because dictionaries and XML-based technologies seem like such a natural fit doesn’t mean that we should not think about what gets left out by language models derived from structured data. A more detailed exploration of these and similar issues could not only help us gain a better understanding of the early trajectories of our field but also perhaps shed new light on the problems we mistakenly considered solved.

DATA MODELING AS AN INTERPRETATIVE FRAMEWORK

3.1 Why model in the first place?

TEI (Text Encoding Initiative) is a set of community-generated guidelines based on an XML schema for marking up text in such a way as to render explicit what is implicit in the structure and the meaning of the text, so that it can be processed reliably by computers or used as an interchange format (Consortium 2007). Text encoding is a major part in the process of designing and publishing a digital edition, but it is not the alpha and omega of it all. A text encoded in TEI is really just that: a text that has been semantically marked up to the degree deemed necessary by its editors. The markup itself may be of interest to those who care about editorial practices and markup theory, but it is hardly an end-product that all scholars and general users can use and interact with. TEI — and text encoding in general — provides an analytic framework for *processing* and *interpreting* data. A digital edition, on the other hand, is a multilayered technical object consisting of the source, the output and the tools to display it (Pierazzo 2011).⁴⁷ The process of creating a digital

⁴⁷ For instance, in the Jane Austen Digital Editions, users can choose whether to switch

edition is a complex endeavor which “requires skills rarely if ever found in any one person” (Shillingsburg 2006: 94). Admittedly, not all electronic dictionaries use TEI — some choose to develop their own models in XML (Müller-Spitzer 2011). Nonetheless, TEI has been a *de facto* standard for text encoding in digital humanities and a large number of retrodigitized lexicographic projects have embraced the format. In Chapter 2, I already outlined the complexity of the dictionary as a text-bearing object. Modeling historical dictionaries in TEI and preparing high-quality digital editions is possible only if we acknowledge, reflect and act upon that complexity in the new medium.

The goal of this chapter is to contextualize the modeling of lexical data within the larger process of preparing digital editions of legacy dictionaries. Unlike Teehan and Keating (2010b), for instance, who discredit TEI as overly complex and unfit for modeling functional, interactive editions, I argue for the primacy of semantic over functional encoding and show how semantic markup is a necessary precondition for interactive editions of legacy dictionaries. By relegating interactivity to an Application Programming Interface (API) with direct access to the semantically marked-up dictionary text, a digital edition built on top of the underlying API can guarantee advanced functionalities without sacrificing the integrity of the data — or the documentary, historical nature of legacy dictionaries.

In previous chapters I have focused on the dictionary itself, as a socio-technical assemblage. I showed how the complex (hyper)textual properties of dictionaries were the basis of rich cultural interpretations and appropriations; and

on or off abbreviations, line breaks, deletions etc. These are all encoded in the TEI *source*, but the user’s interaction with them is made possible in the *output* which consists of JavaScript and XHTML. The edition uses XSLT style sheet as the *tools* necessary to create the interactive output from the TEI source. See <http://www.janeausten.ac.uk/edition/technical.html>

how this ordered yet still chaotic complexity played an important role in both literary imagination and nation building. And while the two planes of this assemblage (the social, cultural and symbolic, on the one hand; and the technical, structural and formalistic), were and are indissolubly linked, I will be now shifting focus to the more formal questions of modeling dictionaries. I speak of modeling, as opposed to simply digitizing, dictionaries because the central problem that all editors of legacy dictionaries face is not how to turn books into their electronic surrogates, but how to represent complex lexical *data* contained in those analogue sources. In this chapter, I will address the questions of modeling and functionality from a largely theoretical standpoint. But this topic will be further fleshed out in the subsequent chapters, especially in Chapter 5, which deals with the particularities of encoding *Lexicon Serbico-Germanico-Latinum* in TEI, and in Chapter 6, which will explore the notion of the dictionary as a service and the digital edition as an API-based interface.

A model is a selective representation of an object or a process with an essentially epistemological goal: to use “a likeness to gain knowledge of its original” (McCarty 2007: 392). Following Geertz (1973), however, McCarty distinguishes between two different types of models: *models-of* and *models-for*:

A model *of* something is an exploratory device, a more or less ‘poor substitute’ for the real thing (Groenewold 1961:98). We build such *models-of* because the object of study is inaccessible or intractable, like poetry or subatomic whatever-they-are. In contrast a model *for* something is a design, exemplary ideal, archetype or other guiding preconception. Thus we construct a model of an airplane in order to see how it works; we design a model for an airplane to guide its construction. (McCarty 2002: 393)

The two types of models have different trajectories: the former takes an existing object and creates its likeness, whereas the latter creates a vision of an object that is yet to be created. One is an interpretation, the other — a projection. In either case, a model is an abstraction from the object it represents: it can never be equal to the object itself.

TEI documents can serve as both *models-of* and *models-for*. A TEI transcription of a hand-written manuscript, for instance, is a semantic model *of* the contents of the original document. It would be very hard to mistake an electronic transcription of hand-written document for the original. But the scope of contextual information and metadata that can be added to a text in a TEI document is potentially infinite and really only limited by the encoder's level of expertise as well as his or her imagination (Driscoll 2006). A TEI document is a model *of* its primary source and a model *for* its future digital edition. It inhabits a peculiarly asymptotic space: it is, potentially, an infinitely expandable receptacle of interpretative data about the source, that — despite the amount of information it can accumulate — will never replace or live up to the materiality of the original. That's why text encoding is an exercise in selection, and one central question for the process of text encoding is always: "where to stop?" (Pierazzo 2011). At the same time, however, even though a model may not live up to the original, it may, in fact, in a certain sense, surpass it. This is what McCarty calls "meaningful failure." The model comes close but does not (and cannot) duplicate the behavior of the primary object: "Its failure as an artifact of engineering is its success as an instrument of science" (McCarty 2002: 371).

The functionality of a digital edition is contingent upon its purpose. But that means that any document — an accounting book, a collection of cooking

recipes or a potpourri of maudlin sentimental poetry alike — has to be functional in the sense that it has to enable the user to perform certain actions on and with the text within a software environment: performing calculations, searching for ingredients or making concordances, depending on the (objective) type of the document and the (subjective) decisions of the editor. A digital edition, like any tool, has to make its *affordances* visible and understandable to the user. The level of detail at which any given text will be encoded is directly proportional to the functionality that the editor envisions for the final product:

the parameters for choosing the level of transcription that is appropriate to a specific project shows that the principal drive behind the decision-making process should be the purpose of the edition. (Pierazzo 2011: 475)

Pierazzo is speaking about diplomatic transcriptions in what she calls the “documentary digital editions” but her conclusions, can be applied to any TEI-based project. The goal of a digital edition should be:

the recording of as many features of the original document as are considered meaningful by the editors, displayed in all the ways the editors consider useful for the readers, including all the tools necessary to achieve such a purpose.

The encoding of legacy dictionaries is no different. My approach to modeling dictionaries is based on the assumption that the field of lexicon retrodigitization should shift — just as it did in the broader field of digital libraries — from quantity and efficiency to quality, precision and user-friendliness; from breadth of electronic content and simple access tools to “sophisticated systems for ongoing use or apparatus providing interpretation” (Lynch 2002). The need to analyze, index and mark-up raw data and provide various types of annotations and metadata, named

entities and other contextual information is of paramount importance for effective searching and retrieval of cultural heritage content (Borin *et al.* ; Borin *et al.* 2007; Schreiber *et al.* 2008; Christopher 2011). When modeling dictionaries, the user's access paths and potential research questions should therefore determine how individual units of dictionary information are encoded (see Gloning and Welter 2001; Müller-Spitzer 2005). This is the case because modeling is, first and foremost, a scholarly challenge, not a technical one. We model in order to first *understand*, then *represent*, and only then *disseminate*. In very pragmatic, project-based terms, this means that the decision what to encode explicitly and what not to encode at all will ultimately be based on the editor's cost-benefit calculation. I will speak more concretely about this process in Chapter 5 *Encoding LSG* (pp. 148-216).

3.2 Functional encoding

Drawing inspiration from Kuutti's (1995) and Nardi's (1995) applications of Activity Theory (AT) to human-computer interaction, Teehan and Keating (2010b) have argued for a modeling approach to historical documents that would encompass not only the semantic structure of the text but also its possible use cases. In this approach, text encoding becomes a (sub)component of a software-engineering process, the end product of which is not simply a tagged text but an interactive document open to user manipulation within a software environment. Consequently, use case modeling is seen as an essential prerequisite for determining the document-specific tagset and schema of an encoded text.

Even though the title of Teehan and Keating's contribution in the *Journal of Linguistic and Literary Computing* — "Appropriate Use Case Modeling for Humanities

Documents” — suggests that they are interested in developing a general modeling approach for digital humanities, the paper refers in fact only to a specialized and never formally defined type of “functional” documents, based on their experience of creating a digital edition of the 18th-century Spanish Alcalà Account Book (see also Keating *et al.* 2010; Teehan and Keating 2010a).⁴⁸ At the same time, however, the authors criticize in broadest terms possible the Text Encoding Initiative (Ide and Véronis 1995; Consortium 2007) as a modeling approach based on a predefined, albeit customizable, schema which is “overly proscriptive” (388) and which can only “sometimes be made to apply in unforeseen circumstances by coincidence, abuse, extension or customization” (385). Because it focuses on text encoding apart from software engineering concerns, the TEI, according to Teehan and Keating, creates documents that are difficult and sometimes impossible to implement in interactive software environments “without significant recoding” (389).

In the rest of this chapter, I will discuss the notion of a functional document from both the software engineering and digital humanities perspectives. I will also analyze whether developing document-specific data models is indeed preferable to applying the existing TEI document data model to humanities documents; how and if the proposed use case modeling is different from “user-driven” TEI encoding (as exemplified, for instance by Rehbein 2010); and what effects document-specific tagsets and schemas may have on interchange and reuse of humanities data.

Teehan and Keating (2010a) are right to point out that a strict division of labor is inscribed in the workflow of most TEI-based digital projects. Humanities

⁴⁸ They base their argument on the Alcalá Account Book manuscript, for which they have created a digital version which not only displays the transcribed text (both in Spanish and English) along with the facsimiles, but also perform actual calculations and answer questions such as “How much was spent on bread at the college in 1788” (383).

researchers and domain experts are taught to encode texts but not necessarily how to create software. Consequently, an encoded text is often “handed off to a software engineer who independently designs a software environment to manipulate it” (381). In their work, Teehan and Keating propose a forward-engineering approach from the object-oriented Unified Modeling Language (UML) that “provides both the software environment and the encoding model” (Ibid.: 98). It is a modeling paradigm that prepares a document — from the very beginning of the text-encoding process — for its inclusion in a software environment. The text is not modeled on the basis of its semantics alone, but rather using logical, physical and interaction classes: the logical class corresponds to the segmented structure of the text itself; the physical model corresponds to the features of the actual artifact that is being digitized (the folios of the manuscript); and the interaction class models possible user interactions with the contents of the digital edition.

Teehan and Keating (2010b) claim that their modeling methodology is appropriate for “historical, functional” documents (381), without fully explaining what they mean by these terms. In the same paper, they also refer to this type of document as “transactional.”⁴⁹ They give examples of what these documents are - “medical records, bank statements and police reports” (2010a: 90), but fail to provide a working definition of such documents. They do mention, however, that there is a difference between between *creative* and *functional* documents (Keating *et al.* 2010) and that “creative humanities objects, such as novels, do not have a definite purpose” and are therefore “less definable” (2010a: 388). The closest we come to a definition of a functional document is that it is a document which “by definition,

⁴⁹ “Transactional documents, that is functional documents or records, are always created in some context and can thus be understood” (Ibid.: 388).

encapsulates functionality and has a number of predetermined uses that are linked to their primary Use Cases.” (51), but this is far from a useful definition because it is entirely circular: saying that a functional document encapsulates its functionality is as informative as saying that a text by default encapsulates its own textuality. How can we establish a clear line separating functional from (supposedly) non-functional (“creative”) documents?

I don’t think that such a line actually exists. It is important to note that Teehan and Keating speak of *documents* rather than *texts*. Documents are concrete, material realizations of texts, which are themselves abstract containers of meaning. Documents are *physical manifestations* of texts.⁵⁰ As physical manifestations, documents are objects. Whether analogue or digital, these objects are used (read, searched on, annotated etc.) by their users. As objects, all documents *afford* certain kinds of use.

The very notion of functional documents is a misnomer because it implies that some documents are functional and others are not. Yet functional documents are, in effect, all documents which exhibit *functional affordances*, i.e. documents that can be purposefully acted upon by their users (see Markus and Silver’s definition of functional affordances on p. 72). Digital editions of novels and other “creative” texts cannot be consumed by their users outside the functionalities of their digital

⁵⁰ It may be helpful to recall here a conceptual model of bibliographic records that’s widely used in library science and which distinguishes between a *work* as a distinct intellectual or artistic creation, its *expression* or “the specific intellectual or artistic form that a work takes each time it is ‘realized’” and *manifestation* or “the physical embodiment of an expression of a work[, which] represents all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form.” (IFLA 1998) Documents in the sense in which they are discussed above are *manifestations*, i.e. physical embodiments of a work that are prepared and delivered as digital editions.

environments. It is not the case that, as Teehan and Keating claim, functional documents “by definition” encapsulate their functionality, but rather that digital editions, by definition, need to be functional.

Furthermore, what Teehan and Keating see as their original contribution to the way text-encoding should be seen as an integral part of conceptualizing a digital edition within a software environment, based on primary and secondary use cases, is really a *de facto* standard of all successful TEI-based projects: defining the project scope and goal, determining the basis for the encoding principles as well as designing a concept for the end product: “its functionality, its features, its navigation” (Rehbein 2010: 4) The TEI Guidelines suggest how to mark up textual segments at various level of granularity and from different analytical perspectives, but they do not include a mechanism for directly encoding potential user interactions with the text or, for that matter, modeling use-case scenarios. TEI is text-centric and software-agnostic. And for good reason, too.

One concrete TEI project that Teehan and Keating criticize is a digital repository of “The Chymistry of Isaac Newton” — a keyword searchable collection of diplomatic transcriptions, normalizations and correlated facsimile images for many of Newton’s manuscripts, including his lab notes (Newman, ed. 2005). According to Teehan and Keating, “the encoding does not provide for the functionality that one would initially expect of such a collection, nor can this functionality be added later, without significant recoding” (2010b: 389). The “significance” or difficulty of recoding is, needless to say, in the eye of the encoder.

The project in question uses the traditional TEI tag set, which does not include labnote-specific tags for “obvious logical model elements such as ‘experiment’, ‘apparatus’, ‘chemical’, ‘method’, or ‘conclusion’” (389). This leads the

authors to the conclusion that it would not be possible to answer queries such as “‘how many experiments did Newton conduct using the chemical copper?’” and that “the ‘comprehensive interactive tools’ that the creators envisage will prove difficult to implement, being conceived of as an additional, rather than an integral part of the digitization ” (389).

This line of argument is faulty on two levels: in terms of assessing the digitization workflow and in terms of what is possible or not within the TEI framework. In their discussion of the use cases, the authors have themselves proposed an iterative (as opposed to sequential) design process (388) which involves “steady refinement of the design based on user testing and other evaluation methods” (Nielsen 1993: 32). Yet the authors don’t seem to recognize the fact that the same cyclical approach of analyzing the requirements, prototyping, evaluating and improving the design can be and is indeed often applied in TEI projects as well. The choice of explicit semantic values that are encoded in a TEI document is always and only an editorial decision. It would be perfectly valid to criticize the editors of the digital collection of Isaac Newton’s “Chimistry” for choosing not to explicitly encode the chemicals or mark up the conclusions of a lab experiment in a separate element. But to claim that the TEI would not be capable of providing a toolset for such mark-up or that it would be extremely difficult to create an interactive digital edition based on the TEI is by no means a foregone conclusion.

There are at least three ways in which a TEI project could accomplish what Teehan and Keating describe as impossible: (1) by using — in a consistent and well-documented manner — a set of attributes with existing TEI elements;⁵¹ (2) by

⁵¹ Take, for example, the following sentence from one of Newton’s lab books: “Camphire dissolved in well deflegmed spirit of niter will make a colourlesse solution.” One could mark-up

extending (and documenting) the TEI schema with elements such as <experiment>, <chemical>, <conclusion>; and (3) by integrating a controlled vocabulary for chemical elements within the existing TEI framework.⁵² In all of these cases — as long as the encoding is consistent and well-documented, creating a mechanism for answering the question “What experiments used such and such a chemical” would be trivial using XPath, a standard query language for selecting parts of an XML document (Kay 2008), or XQuery, a functional programming language (Walmsley 2015).

When all is said and done, the problem with Teehan and Keating’s approach is the problem of perspective, priority and privilege. I argued above that the functionality of the document is contingent upon the purpose we ascribe to it as editors or readers. Teehan and Keating’s approach, on the other hand, stems from the perspective of a software engineer. In their papers, Teehan and Keating suggest (a) that efficiency is an important aspect in the software development process; and (b) that the source and the tools created to display and manipulate the source need

camphire and niter with the TEI element <term>, which, according to the Guidelines, “contains a single-word, multi-word, or symbolic designation which is regarded as a technical term” (Consortium 2007). To further specify the nature of the term, we could add the attribute value “chemicalCompound” to it: <term type=“chemicalCompound”>Camphire</term>. It would be a trivial matter to write, for instance, an xQuery script to retrieve all experiments containing camphor.

⁵² One could do that in two ways: (a) by using an external file listing all (possibly normalized) chemical elements and compounds that appear in Newton’s texts and then referencing the appearances of chemical elements in the text with the external file using the referencing string <rs>, a TEI element that contains “a general purpose name or referring string.” In that case, we would have something like: <rs ref=“compounds.xml#camphor”>Camphire</rs> where the ref attribute would be pointing to a unique element in the reference list; or (b) by pointing, in a similar fashion, to a unique entry in the existing registry of chemical compounds such as the CAS (Chemical Abstracts Service) (see <http://www.cas.org/expertise/cascontent/registry/index.html>) or the Registry or the Royal Society of Chemistry’s free database ChemSpider (<http://www.chemspider.com/>).

to be intricately connected (again, with the goal of making the end product as flexible and as easy to produce). Theirs is a valid point — from the perspective of the software engineer tasked with the goal of creating a software environment for an interactive digital edition. But their approach privileges the role of the software engineer above the other essential stakeholder — the humanist scholar, whose initial priority is to understand the text and only then, based on the understanding of the text, make it available in an interactive environment. Having a fully-developed standardized output at an early stage of text encoding of the work may indeed be premature and “misleading... on how to do things” as Rehbein (2010) suggests. Text encoding is, among other things, a subtle and detailed process of getting to know the text.

Teehan and Keating’s approach does not in any way help with the crucial problem of divided expertise between humanists and software engineers. The claim that they are interested in creating “Human Usable Documents for wide community use” (386), but they are really instrumentalizing the notion of the document itself. The same way that they accuse TEI of being proscriptive about the schema, they are proscriptive about what they see as primary use cases. The three-class document model may make things easier for the engineers, but not necessarily for the domain experts: “the humanities researcher who encodes XML must also be aware of the interaction class and the requirements of the software engineer.”(Teehan and Keating 2010a: 387).

And even though they base their approach on various use cases, which is necessary in any digital project, their approach also limits the extent to which source documents can be repurposed for use cases other than the ones originally envisioned. The marriage of text encoding to engineering classes (or, in other

words, of the source to the tools that create its output) may be a happy union at first, but it is a marriage of convenience, and as such not necessarily well-prepared to withstand the test of time, as contexts change, technologies evolve, and individual perspectives shift. Most importantly, the coupling of text encoding and engineering production is not conducive to interoperability and successful integration into a larger family of resources. It's a hit-and-miss affair.

It is beyond the scope of this chapter to address the question of interoperability in great depths. Suffice it to say, TEI is *not* an out-of-box solution for interoperable textual resources. In fact, TEI is a morally ambiguous creature because its main virtue is also its main vice: it is flexible to the point of fatigue. When a single textual entity can be encoded in various ways, the dream of straight-forward, automatic, unmediated integration of different textual resources has very little chance of becoming a reality. Even if there existed a one-to-one correspondence between textual phenomena and ways of encoding them, the epistemological problem would remain: models have a tendency to decontextualize and de-historicize. Having said all that, however, a common vocabulary such as the one provided by TEI and a common set of instructions such as those provided by the *TEI Guidelines* can go a long way at creating an ecosystem of textual resources that share the same editing principles and that can be, if need arises, integrated. The integration of textual resources, however, is not so much a task for TEI itself, which is an encoding standard, but for the emerging digital infrastructures such as CLARIN or DARIAH.

Teehan and Keating are right to point out that “it seems overly proscriptive to imply that this [TEI] is the only valid approach, and that it encapsulates all possible uses and perspectives” (386) — only it is not clear whether anybody ever

implied let alone claimed in all seriousness that TEI does encapsulate all possible uses and perspectives. The TEI community is certainly not lacking in big egos, but madness is generally well-contained.⁵³

In subsequent chapters, I will show very concretely how an interactive, exploratory digital edition of a historical dictionary can be built within a system architecture which is based on a clear separation of roles: a TEI-encoded source which provides structured data; an API which provides direct access to data; and an online interface (or digital edition proper) which serves as an interface between the user and the data.

⁵³ There is one aspect of Teehan's and Keating's criticism of TEI that might be worth exploring in the future: "We believe that the XML/TEI approach should at least be broadened to support the entire project life-cycle. Currently the guidelines foreground the use of XML/TEI as a solution, in isolation from accompanying software, and before the problems of the project are even enunciated." (ibid.: 92) Even though modeling use cases and the software production cycle goes beyond the scope of TEI as it stand now, one could entertain the idea of how to make that happen, perhaps as part of a future chapter on modeling born-digital resources that were not created in TEI, or by considering a possible interface between text encoding and critical code studies.

LEXICON SERBICO-GERMANICO-LATINUM

4.1 A literary language for a ‘bookless’ people?

In 1815, German philologist Jacob Grimm — future author, together with his brother Wilhelm, of the monumental *Deutsches Wörterbuch* — wrote a review of an anthology of Serbian folk poetry that had been published the year before by Vuk Stefanović Karadžić, a young Serbian scholar living in Vienna. The article appeared in the *Wiener Allgemeine Literarische Zeitung* (Grimm 1815/1869). In it, Grimm praised the popular Serbian songs for their freshness and natural beauty, their prosody and soulfulness. Above all, Grimm put the Serbs as a kind of collective imaginary on the pedestal of folkloric, “bookless” authenticity:

Unter allen slawischen völkerstämmen sind diese Serben, mit ihrer sanften, überaus singbaren sprache, zum voraus begabt mit lied, gesang und sage, und es scheint, als ob der gütige himmel ihnen ihre bücherlosigkeit durch einen hausseggen von volkspoesie stets habe ersetzen wollen. (436)

[Of all the Slavonic peoples, the Serbs are by virtue of their subtle and melodic language the most gifted in poems, songs and stories, and it

looks as if the good God had by this rich blessing of popular poetry wished to make up to them for their lack of books.]

Grimm's view of the Serbs — and other Slavic nations — as the singing savages of Europe was not particularly novel at the time. As an expression of the pre-Romantic infatuation with the “noble savage” (see Ellingson 2001), it combined philological evidence of ethnolinguistic diversity with an organicist interpretation of language as a reflection of the national character of a people. Herder's *Abhandlung über den Ursprung der Sprache* [Treatise on the Origin of Language] (1772) identified language as the quintessential human characteristic, and *das Volk* (the people) as a quintessentially linguistic community.⁵⁴ His *Philosophie der Geschichte der Menschheit* (1784-91) and his two-volume collection of folk songs *Stimmen der Völker* (1778-9) solidified the view that the study of foreign cultures through the conduit of language can reveal their immanent poetic and philosophic realities.

Needless to say, Grimm's characterization of the Serbs as a tribe “lacking in books” masked the complexities of the linguistic and literary divisions in the Serbian cultural sphere at the beginning of the 19th century. The majority of Serbs, especially those living under Ottoman rule, were still illiterate, while the lion's share of the literary production came from the Serbian centers of learning in the Habsburg Empire. In the absence of a national state that could form institutions necessary to create a binding literary standard, and with the population split between two empires, the literary production of Serbian writers was dominated throughout the 18th century — to various degrees at various junctures — by three different linguistic registers: (1) the ecclesiastical Serbian recension of Church

⁵⁴ A human being deprived of language would be “das traurigste sinnloseste verlassenste Geschöpf der Schöpfung; und der größte Widerspruch mit sich selbst” Werke 1:776.

Slavonic (Serbo-Slavonic); (2) the Russified recension of Church Slavonic known as Russo-Slavonic; and (3) a fashionable hybrid of Russo-Slavonic and vernacular Serbian known as Slaveno-Serbian (Albin 1970; Младеновић 1989; Kretschmer 1994, 2004, 2008; Кречмер 2007; Албијанић 2010; Ilić Marković *et al.* 2012).

By the time Vuk Stefanović Karadžić had arrived on the Serbian literary scene with the publication of a first vernacular grammar — *Писменица сербскоја језика* (1814) and his first collection of Serbian folk poetry, the more archaic Serbo-Slavonic and the Russo-Slavonic hybrid were already too removed from the vernacular and used by far too few authors to be considered a viable option for language standardization. Slaveno-Serbian, however, enjoyed significant prestige: it was considered to be a ‘true’ literary language as opposed to the ‘simple’ and ‘vulgar’ language of the peasants.

Throughout the 19th century, a venomous debate on the direction of the literary language had consumed the Serbian literary scene with arguments and insults alike being hurled from both sides of the linguistic divide. With speakers of Serbian living in two empires (the Ottoman and the Hapsburg) and with the majority of secular cultural production taking place in the Hapsburg Vojvodina (Скерлић 1966; Ивић 1971), Slaveno-Serbian was championed by the Serbian Orthodox Church as a political tool for preserving cultural ties with Russia and resisting Austrian attempts to introduce the Latin alphabet and spread Catholicism among the Serbs (Добрашиновић *et al.* 1980; Grčević 2009).

The Serbian linguistic landscape at the beginning of the 19th century could be described as essentially *diglossic* (Ferguson 1959), in the sense that two language varieties, the vernacular Serbian („народни језик”) and the literary Slaveno-Serbian („славено-српски”), performed functionally complementing roles within

the same linguistic community: the former was largely the vehicle of everyday communication, corresponding to Ferguson's "low variety" (L), while the latter, as a "high variety" (H) was used mostly for written and formal purposes. The two varieties differed in terms of prestige, literary heritage, acquisition, orthographic and grammatical complexity and lexicon, to name just some of Ferguson's distinguishing features of diglossia. At the same time, however, an important distinction needs to be made in terms of standardization and stability. In the cases of typical diglossic languages, such as classical and vernacular Arabic, High German and Swiss German, or Katharevousa and demotic Greek, for instance:

there is a strong tradition of grammatical study of the H form of the language. There are grammars, dictionaries, treatises on pronunciation, style, and so on. There is an established norm for pronunciation, grammar, and vocabulary which allows variation only within certain limits. The orthography is well established and has little variation. (Ibid.: 331-2)

This was definitively not the case with Slaveno-Serbian. In the absence of established standardization bodies or widely accepted grammars, dictionaries and style guides, Slaveno-Serbian was a constantly moving target. Books were being written, published and read in Slaveno-Serbian, but with little orthographic and grammatical consistency: writers developing their own idiolects based on their native dialect, familiarity with Russian or Serbian recensions of Old Church Slavic and a fair amount of personal whim. This has led previous generations of scholars to describe Slaveno-Serbian as an "artificial language" (Albin 1970: 490) and as „[нешто] што није језик већ језичко стање" ("not a language but a linguistic condition" Грицкат 1966: 65), which in view of the more contemporary explorations of hybridity and mixed languages (Bakker and Mous 1994; Matras and Bakker 2003)

could be considered to be both a typological and an ideological oversimplification. Slaveno-Serbian served as a conduit of intellectual and artistic exchange from the Baroque to the Enlightenment, and Slaveno-Serbian writers had a wide array of stylistic devices at their disposal (Иванова 2010), but the fact of the matter remains that — unlike a typical H language — Slaveno-Serbian was neither inherently stable nor highly codified (Кречмер 1988; Младеновић 1989).

In these matters, Vuk⁵⁵ was strongly influenced, both ideologically and practically by Jernej Kopitar, a Slovenian scholar working for the Imperial Library in Vienna as a censor for Slavic books (Butler 1969; Grčević 2009). In the *Одъявление о Сердскоме Рѣчнику* [Announcement of the Serbian Dictionary], which was published as an unpaginated insert to issues 88, 112 and 172 the Viennese *Новине сербске* [Serbian News] in 1816, he raises the question of what kind of dictionary was needed. His answer is unequivocal: „требамо найпре Рѣчникъ народноѣа Сердскоѣа языка, у коемъ ће се содержавати само оне рѣчи, кое народъ Сердскѣй употребљава” [we need above all a *Dictionary of the Vernacular Serbian Language*, which will contain only those words that the Serbian people use]. Vis-a-vis Slaveno-Serbian, Vuk seems to be conciliatory at first:

Славенскѣй языкъ остае источникъ обогащенѣа Сердскомъ языку, као и Россѣйскомъ. Но Рѣчникъ Сердскѣй мора (садъ за первый кратъ) представити Сердлемъ нѣиовъ собственнѣй языкъ — најсветѣю народность — онако какавъ онъ самъ по себи есть. Онъ мора изъ са-

⁵⁵ In Serbian public discourse, it is customary to refer to Vuk Stefanović Karadžić by his first name. Furthermore, non-Serbian scholars have carried on with this peculiar tradition in their own languages. I have argued before that this is a case of false familiarity: by calling the reformer of the Serbian language by his first name, we are evoking intimacy with somebody who is, in fact, foreign to us (Тасовац 2007). I am adopting this naming convention in English quite reluctantly and only for the sake of simplicity.

мог' народа, кои нѣмъ говори, почерпанъ быти. (Караџић 1816/1969)

[The Slavonic language remains a source of enrichment for the Serbian language. But a Serbian Dictionary must (in the first instance) present to the Serbs their own language — the holiest nationality — as it is on its own. It has to be extracted from the people themselves who speak it.]

Initially, Vuk claimed that Slavonic words used by the “noble and educated part of the people” and those used „за прїятїе богатства и красоте чужег' кнѣжества” [for accepting the riches and beauties of foreign literatures] will need a separate dictionary, as an addition to the Serbian dictionary proper. Between the two dictionaries one would imagine „тѣсанъ, лакъ и прїятанъ союзъ” [a close, easy and pleasant association] (Ibid.). In the Second Announcement (Ibid.), there was already no more talk of a Slavonic Dictionary, with Vuk increasingly focusing his attention on the criticism of Slaveno-Serbian, insisting that the situation in which writers compose their works in a language “without any rules” and “each according to his own taste” is untenable (Караџић 1818).

Even though Vuk was not the first proponent of the literary language based on the vernacular,⁵⁶ he became over time its most vocal — and most controversial — champion. Throughout his life, he continued to collect and publish folk poetry and prose, and in 1847, he also published a vernacular translation of *The New Testament*. By 1850, when Vuk and a group of Serbian and Croatian writers signed the so-called Vienna Literary Agreement, which programmatically established a joint literary language for Croats and Serbs (Auty 1958; Dronjic 2011; Greenberg 2011), his

⁵⁶ a simplified Serbian alphabet had already been worked out by Sava Mrkalj (see Ilić Marković *et al.* 2012).

linguistic reform — with all the political implications involved therein — was complete.

4.2 Typology of the dictionary

Serbian scholars generally tend to refer to the dictionary by its title (“Serbian Dictionary”) almost always leaving out the subtitle (“explicated in German and Latin”). This is, on the one hand, a matter of convenience, but I do believe that this kind of shorthand, while saving some space in scholarly articles, occludes the scope and structure of the work *and* in doing so indirectly misrepresents its typological peculiarities.

Vuk’s *Serbian Dictionary* is *not* a monolingual dictionary of the Serbian language. That is why I have chosen, throughout this dissertation, to refer to the dictionary by its Latin title, as *Lexicon Serbico-Germanico-Latinum* or *LSGL*. Typologically, *LSGL* displays some features of a bilingualized dictionary (Béjoint 1994), a dictionary whose primary language of explication is Serbian (the lemma, the definition) but with lexical equivalents in a different language. Considering that lexical equivalents in *LSGL* are provided in both German and Latin, it would perhaps be more appropriate to call this dictionary trilingualized, but this term is not used in the scientific literature. The difficulty here, however, is not only terminological but statistical. *LSGL* in E5 has 46,967 entries, of which only 5302 contain lexicographic definitions in Serbian, while 1043 entries in total also contain additional explicatory narratives. All together, 5939 entries contain either monolingual definitions *or* narrative sections in Serbian. This means that only some 12% of the dictionary contain significant monolingual components beyond

examples. Of the 46,967 entries, 29,935 have translation equivalents in German and 29,180 in Latin. 16,773 entries provide no translation equivalents at all. Of these, 15,074 are cross-references to other entries in the dictionary and 1,699 have only monolingual content (either definitions, examples or both). These numbers speak for themselves: SLSG is primarily a bilingual dictionary with some bilingualized, i.e. trilingualized content. At the same time, however, SLSG is not a typical bilingual dictionary either because, for instance, examples, of which there are altogether 12,486 in 9,274 entries, are *not* glossed in either German or Latin.

This, however, is not the only typological curiosity related to this dictionary. Because Vuk's dictionary is usually seen as part of a revolutionary narrative — a narrative of rupture *and* rapture — its dialectological span is rarely mentioned when discussing its broader cultural relevance. The dictionary is often hailed for its broad, encyclopedic scope, which it undoubtedly has,⁵⁷ but the spotlight is rarely cast, especially among the general public, on fact that the “Serbian Dictionary” is also a dictionary of Serbian dialects.⁵⁸ Here, too, however, numbers speak for themselves. Of the 46,967 entries in the dictionary, 13,303 have diatopic usage labels (see Section 5.3.5.2 [pp. 166-167]). In other words, almost 30 percent of the entries can be characterized as regionalisms. Vuk was very much aware of this fact. This is why, in

⁵⁷ „Вуков Рјечник није обично лексикографско дело, него нешто много више од тога, енциклопедија српског народног живота” [Vuk's *Dictionary* is not a typical lexicographic work, but something more than that: an encyclopedia of Serbian folklore (lit. “national life”)] (Деретић 1990)

⁵⁸ For instance, the Wikipedia article on *LSGL* describes it as „један од најважнијих корака у борби за језик јер је постављао за основу књижевног језика чисти народни језик” [one of the most important steps in [Vuk's] linguistic fight because it set as the basis of the literary language — pure vernacular (lit. national/people's language)]. Leaving aside the rhetoric of linguistic purity, which often accompanies texts about Vuk, statements like this often imply that the vernacular is a single unity.

his preface to the Second Edition, he stressed that the multiplicity of dialects represented in the dictionary could yet not amount to a single literary standard, but that until such a standard was eventually adopted, based on one such dialect, the main goal was to become familiar with „народни језик по свијем крајевима” [the vernacular in all the regions] (Караџић 1852a: unpaginated).⁵⁹

4.3 Importance of the dictionary

As the first lexicon of the modern Serbian vernacular(s), Vuk Stefanović Karadžić's *Српски рјечник исцумачен њемачкијем и латинскијем ријечима. Lexicon Serbico-Germanico-Latinum* (1818, 1852) has a unique place in the history of not only the Serbian language, but the South Slavic diasystem in general (Дмитриев and Сафронов 1984; Wilson 1986; Стојановић 1987; Eschker 1988; Potthoff 1990; Bockholt 1990; Ивић 1990; Vitalich 2005; Кулаковский 2005). The *Lexicon* had a profound influence on the development of South Slavic lexicography following the above-mentioned Vienna Literary Agreement. It left indelible traces on the major dictionaries that followed well into the twentieth century, such as the two-volume *Dictionary of the Croatian Language* by Iveković and Broz (1901), the 23-volume *Dictionary of the Croatian or Serbian Language* by the Yugoslav Academy of Arts and Sciences (JAZU 1880), the six-volume *Dictionary of the Serbocroatian Language* by Matica Srpska (MC 1967), as well as the ambitious but still incomplete 30-plus-volume *Dictionary of the Serbocroatian Literary and Vernacular Language* by the Serbian Academy of Arts and Sciences (САНУ 1959).

⁵⁹ See the full quote on p. 188.

For the scholarly community, the *Lexicon* has been a major resource for etymological research. Examples include Petar Skok's classic *Etymological Dictionary of Croatian or Serbian Language* (Skok 1971) and the more recent *Etymological Dictionary of the Serbian Language* — an ongoing project of the Institute for the Serbian Language of the Serbian Academy of Arts and Sciences (Бјелетић *et al.* 2003). What's more, scholarly interest in the genesis, evolution, standardization and disintegration of South Slavic languages has increased considerably in recent decades, as attested by various studies including Bugarski and Hawkesworth 1992; Okuka 1998; Kunzmann-Müller 2000; Magner and Marić 2002.

The dictionary text is rich with ethnographic and anthropological material. Not only do many entries contain examples of Balkan folk storytelling, but some are themselves structured as historical, cultural and ethnographic narratives that offer informative sketches and sometimes even very detailed accounts of the myths and realities of the Balkan past (see, for instance, entries for *кмет*, *ошмица*, *мора*, *хајдук*, etc.). While the majority of lexemes that appear in Karadžić's dictionary come from the Neo-Štokavian/Jekavian dialect⁶⁰ spoken in his native region of Tršić, he collected materials for his dictionary across the Balkans. It is not uncommon to find examples from different dialects and regions, including Dubrovnik, Dalmatia, Eastern Montenegro, Boka Kotorska or Vojvodina.

The literature on Vuk Stefanović Karadžić is too extensive and varied to be given but a cursory overview here. For the purpose of this research, it is important to stress the fact that *The Lexicon* is well-researched, both as a lexicographic and

⁶⁰ On the complex mixture of dialects in the South Slavic *Sprachbund*, see Lukežić 1990; Greenberg 1996; McLennan 1996; Friedman 1997; Белић and Пијановић 1999; Zečević 2000; Белић 2000; Lisac 2003; Ивић 2009.

literary work (Ивић 1966; Поповић 1983; Милинчевић 1987; Ивић 1988). There is a wealth of studies on various aspects of the dictionary, including: definitions (Вукомановић 1976; Станић 1978; Дешић 1996; Вукомановић 1999), grammatical innovations (Тјапко 2011), accentuation (Симић 1981), demonyms (Штасни 2012), personal names (Медведев 1990), Turkisms (Пецо 1986; Пецо 1987b), Germanisms (Пецо 1987a), Slavenisms (Цветковић 2005), and Hellenisms (Слапшак 1987). The phonological, morphological and semantic doublets have been analyzed by Кашић (1973); stylistics by Симић (1988), and pejoratives by Стакић (1988). Other important areas of research included Serbian epic poetry as a source for the dictionary (Матић 1960; Латковић 1964; Младеновић 1966; Матицки 1978; Сувајцић 2011), the reception of the Kosovo myth in the dictionary (Delić 1990), legal terminology (Елаковић-Ненадовић 2011) and the sociolinguistic aspects of the dictionary (Матијашевић 1988). Even culinary recipes from the dictionary have been the subject of scholarly attention (Ивић 2003).

4.4 Different editions

4.4.1 First edition (1818)

Vuk published the First Edition of his dictionary (hereafter, E1) in Vienna, in 1818 (Краџић 1818). E1, which contained 26,270 entries, was to a large extent based on the Jekavian dialect of Vuk's native Tršić (Ивић 1966).

4.4.2 Second edition (1852)

The Second Edition (hereafter E2) was considerably larger than the first edition. In it, Vuk incorporated lexica which he collected on his travels through Croatia, Dalmatia, Dubrovnik, Boka and Montenegro (Краџић 1852a), making the dictionary

much more representative of the South Slavic *Sprachbund*. Significantly, however, E2 did not contain almost any of the much-criticized vulgarisms that were included in E1. Obscene vocabulary was pruned out of the dictionary and would not be included in later, posthumous editions either. The tradition of linguistic purism would carry over into other major Serbian dictionaries well into the twentieth century. In addition to *verba obscena*, some entries were accidentally left out of the Third Edition.

4.4.3 *Incomplete third edition (1892)*

Jovan Bošković took it upon himself to revise and edit a first posthumous edition of L2, but only the first of the planned volumes was finished and published before his death (Караџић 1892). Bošković's envisioned a significant update to the dictionary with etymologies, new words (not used by Vuk himself) and new synonyms.

Ну ово издање не смије бити ни само просто прештампавање. За прошлих 40 година (1852-92) језик је живио и порастао (лексикално) а наука о њему, и оџиџа и џоседна срџска и слоенска баш највише напредовала, изведавши за то вријеме своја главна дјела. Према томе, просто прештампавање значило би, с једне стране, не хтјети знати за науку, а с друге, захтијевати од 30 до 50 тисућа образованих читаца да знају све оно, што може знати стотинак стручних људи. (Ibid.: xiii)

[This edition shouldn't be merely a reprint. Over the past 40 years (1852-92) our language has lived and grown (lexically) and the science of it, both *general* and *particular Serbian and Slavonic* has progressed the most and produced its most important works. A mere reprint would mean that, on the one hand, one doesn't care about science, and, on the other, that 30-50,000 educated people would have to know all that only about hundred scholars can know.]

Bošković decided to keep the original (including smaller changes that Vuk himself planned to introduce) — calling this the “historical dimension” of the Third Edition („[т]аки текст *їредсїавља истїоријску сїрану* трећег издања”, xiii), while his additions were added in square brackets as scholia („схологије, т.ј. критичке и објашњајне напомене”) or glosses („ілосе”, xiii) to “remind us of the developments in language and the progress in science for the last forty years” („іодсјећају на развиїак језика и најредак науке, за прошлих 40 година.)

4.4.4 Third edition (1898)

The complete Third Edition (hereafter E3), also referred to as “the state edition” because it was published after Vuk’s death by the State Printing House of the Kingdom of Serbia (Караџић 1898) was revised by Pera Đorđević and Ljubomir Stojanović, but with a different editorial concept than the one presented by Bošković. The editors of E3 explicitly set out to create an edition that would be corrected and expanded („исїрављено и доїуњено”) but that would remain “Vuk’s own” („да... опет остане Вуково”):

тј. да у речнику исправе и допуне само оно што би и Вук учинио, да је одмах после свога другог спремао за штампу и ново, треће, издање речника. Тога ради, не уносећи у речник нових речи ни из других Вукових списа ни из народних умотворина које је Вук издао, потписани су се непрестано у свему држали II. издања Вукова (1852.) и сада, у овом новом издању. (Караџић 1898, xiii)

[i.e to correct and expand only those things that Vuk himself would have done, had he after his second prepared a new, third edition for publication. For that reason and without adding new words to the dictionary either from Vuk’s other works or from folk literature published by Vuk, the undersigned kept true in everything to Vuk’s 2nd edition (1852) in this new edition.]

E3 included Vuk's marginalia from his own copies of the first two editions (new words, new senses, examples, various types of notes); it fixed a range of typographic mistakes and omissions (although not all, most notably the censored *verba obscena*); and extended cross-references using elaborate typographic conventions to separate their own interventions from the original text (see Section 4.6, pp. 141-145).

4.4.5 *Fourth edition (1935)*

The Fourth Edition (Караџић 1935, hereafter L4, was a facsimile edition of the third edition of a slightly different size and printed on better-quality paper (see „О овоме издању” in Ibid., unpaginated).

4.4.6 *Differences between editions*

It is beyond the scope of this chapter — or this dissertation as a whole — to explore in great detail the genealogy of *LSGL*'s different editions. The changes based on Vuk's own hand-written annotations of E1 and E2 are meticulously described in the Preface to E3 and will be mentioned here only briefly in order to highlight both the level of detail with which the posthumous Third Edition was edited and to set the stage for the following chapter, which will show how these differences have been encoded in the digital *LSGL*.

4.4.6.1 *New content from E2 marginalia*

Some entries which Vuk jotted down on the margins of his personal copy of E2, were incorporated in the macrostructure of E3 but typographically set apart. For instance:

бесѡмучан, чна, чно, (ист.) vide бјесомучан.
бѣспамѣт, f. (у Кучима): Шта учини, беспамети! Особито реку дјевојци која пође за удовца. vide безумље.
бѣспара[*], f. vide [судоперал] опирњача.

The entry *бесѡмечѣ* (lit. lack of reason, but used here figuratively in the sense of mad person) did not exist in E2. It was transferred from Vuk's notes and placed in the alphabetic order between *бесомучан* [raving, frenzied] and *бесѡпара* [dish cloth], along with a diatopic note (у Кучима, [in Kuči]), an example, a usage hint (*Особито реку дјевојци која пође за удовца* [Especially said to a girl who marries a widower]) and a cross-reference. As can be seen from the above image, the marginal entry was printed in italics in order to set it apart from the “regular” entries.

4.4.6.2 Corrected lemmas

In some cases, Vuk noted typographic and other errors in E2. Those were, in the third edition, also marked up using a different italic font. Compare, for instance, entries *орепак* in E2 and E3:

ОРЕПАК, пка, m. das leichte (und schlechte) Getreide, welches beim Worfeln neben das schwere vertragen wird, grana leviora: оснјѣци тај орепак. cf. отражак.

орепак, пка, m. das leichte (und schlechte) Getreide, welches beim Worfeln neben das schwere vertragen wird, grana leviora: оснјѣци тај орепак. cf. отражак.

In E2, the lemma was accented on the first syllable with a long rising accent (*ópenak*). In E3, the accent was changed, following Vuk's correction, to short rising (*òpenak*). See also *бадар, Марко Краљевић, Чајниче, љаљак, лежај, кутлош* etc.

4.4.6.3 Deleted content

Content that was marked up by Vuk for deletion in E2 has been removed from E3 as well. Consider, for instance, *Jě̑ap* in E2:

ЈЕГАР, гра, т. 1) (ст.) некако мјесто:
Ти нијесн друма погодно,
То не нде Јегру бијеломе,
Већ Јањоку граду каурскоме —
2) Stadt Erlau in Ungern, Agria.

The entry has two senses: the first sense is marked as poetic or elevated⁶¹ and defined as *некако мјесто* [some kind of place]. This sense is also accompanied by an example from a folk poem. The second sense contains only translations into German (Stadt Erlau in Ungern [sic!]) and Latin (Agria).

In E3, following Vuk's own notes, the underspecified definition was removed, the entry became monosemous and the example, which was associated with the first sense in E2, became an illustration of the meaning Erlau, Agria.⁶² Curiously enough, the mistake in the German (Ungern instead of Ungarn) has not been corrected.

⁶¹ For an explanation of *cū.* as usage note, see Section 5.3.5 on usage in general (pp. 163–173) and Section 5.3.5.4 on diaphasic labels, in particular (pp. 170–171).

⁶² The interpretation that *Jȇap* and the quoted poem in fact refer to the Hungarian city of Eger (Germ. Erlau, Lat. Agria) has been subsequently confirmed by both (Iveković and Broz 1901: I, 484) and (JAZU 1880: IV, 567)

Јегар, гра, m. Stadt Erlau in Ungern, Agria:
 Ти нијеси друма погодио, То не иде Јегру
 бијеломе, Већ Јањоку граду каурскоме —

4.4.6.4 *Accidental omissions*

E3 reconstructed some of the missing content from E3 that was present in E1, but not *verba obsцена*, which were intentionally removed from E2. The accidental omissions included entries: доцњење, разгрејати, врајт, 2 скуп, троји, etc. They appear in E3 within square brackets along with a note „из I. издања” [from the First Edition].

4.4.6.5 *Marginalia from E1*

Vuk’s marginalia from E1 were added in square brackets with a note „Маргиналија I. издања” [First Edition marginalia]. These included, for instance, entries such as оѡармачиѡи, оѡаѡрнуѡи, ѡресвласчење, the explanation for Вучиѡрн, or narrative additions to entries such as задушнице, крсно име, and женидба.

4.4.6.6 *E3-specific additions*

The editors of the third edition also created a number of entries for words which appear in the body of E2, but not as entries in their own right. These lexemes can be divided in two groups:

1. *Words appearing in dictionary content*, either in definitions proper, commentaries, or examples, but lacking entries of their own. They were, according to the editors, „превођене или објашњиване начином како је и Вук радио у II. издању” (xiv) [translated or explained in the same way in which Vuk did it in the Second Edition]. Such words were cross-referenced with entries from which they were excerpted using the label види s. v. [see

sub voce]. In addition, such excerpted words were printed with extended letter-spacing in the source entries in order to mark their special status. For instance, the lexeme *гpaйoцјeн, a, o* [precious] has been excerpted from the definition of the first sense of the lexeme *заклад* (das Kleinod, ornamentum): „лијепе хаљине, оружје, и остале д р а г о ц ј е н е ствари” because there was no entry for *гpaйoцјeн* in E2. The newly added entry for *гpaйoцјeн* was placed in square brackets to indicate that it has been added in E3.⁶³

2. *Words that appeared only as cross-references but were missing an entry of their own* were added as independent entries in E3 inside square brackets, also referring to the source entry. For instance, the second sense of entry 1 *paшaк* in the E2 contained a cross-reference (vide *кpoпмip*), and the entry *кpйoлa* contained a cross-reference (cf. *кpoпмip*), but the lexeme *кpoймip* did not exist as an entry. In E3, this missing entry was added in square brackets: [кpoпмip, m. vide кpтoлa.]

4.4.6.7 *Cross-references introduced in the third edition*

In addition to creating entries for cross-references without a target, as described in the previous section, the editors of E3 also systematically went through the entire text of E2 and added a number of cross-references that were not present in E2. These were also set apart by square brackets.

4.4.7 *Second Edition Reprint (1986)*

In a curious attempt to secure the “authenticity” of Vuk’s dictionary, the editors of his *Collected Works* decided to reprint E2 in 1986 without the subsequent editorial

⁶³ See also Section 5.4.2 (pp. 204-208).

interventions from E3, but with the inclusion of Vuk's marginalia and other changes introduced in E3 in the notes at the back of the dictionary (Караџић 1986a; Караџић 1986b). This was a highly impractical decision. One of the virtues of the E3 was that it developed a consistent set of typographic conventions to mark up the changes introduced. The edition in the *Collected Works*, however, introduced no new editorial content, but only moved the annotations, which had already been included in the text of E3, to the endnotes of the Second Edition Reprint. In this edition, unnumbered endnotes refer to page numbers of the dictionary, but there are no pointers in the dictionary which would indicate whether entries have been annotated in some way. As a result, the user who consults the dictionary would have to browse the endnotes for each and every entry he or she consults to check whether the given entry was annotated in some way. The process is cumbersome and ultimately frustrating.

4.4.8 Digital Edition (2017)

The digital edition, which is an integral part of this dissertation, has been produced on the basis of the text of E4, a facsimile edition of E3. The implementation of the digital edition consisted of several interconnected phases:

1. *Text Processing (capture, encoding and enrichment)*. The textual content of the dictionary was manually keyed, preserving basic typographical features of the text (bold, italic etc.). Subsequently it was encoded according to the *Guidelines of the Text Encoding Initiative*.
2. *Backend and API Development*. The backend and the API were developed using eXist-db, a native XML database, and xQuery.

3. *Frontend Development.* The frontend was developed in PHP, using the Laravel PHP Framework and JavaScript. The frontend was connected to the backend via the API.

I will discuss the encoding of the dictionary in the next chapter, and focus on the API and frontend development in Chapter 6.

ENCODING *LSGL*

5.1 Encoding principles

The encoding of a legacy dictionary is a process which consists of three mutually related phases:

1. the *analysis* of a concrete dictionary, its content and its typographic conventions;
2. the *mapping* of the concrete dictionary onto the abstract model of dictionary structure;
3. the *translation* of the said model into the TEI-XML serialization.

While it is, in principle, impossible for editors to engage in Phase One without having, at least in the back of their mind, the abstract model of dictionary structure, one should always try to get to know the text, which is to be encoded, *on its own terms* — as much as our overtrained and overly conditioned editorial egos will let us. Once editors have familiarized themselves with a concrete dictionary and its textual conventions, they can proceed to *map* this particular text onto the abstract model of dictionary structure: the *macrostructure* (systematic arrangement of lexical items in

the dictionary); the *microstructure* (systematic arrangement of information *within* a lexical item in a dictionary); and the *mediostructure* (a systematic network connecting various data points in the dictionary). Equipped with a sound understanding of the how the abstract dictionary model is instantiated in the given dictionary text, the editor can proceed to translate the this particular instantiation into a TEI-XML serialization.

It should be emphasized at the outset that TEI is *not* a dictionary-specific serialization. As of Version 3.2.0, updated on July 10th, 2017, TEI had 569 elements to describe various textual genres (manuscripts, drama, verse etc.), entities (names, dates, people, places etc.), data structures (feature structures, graphs, networks and trees) etc. TEI is a general-purpose vocabulary for encoding texts in the humanities. This will have important ramifications for the encoding of dictionary data in TEI. Because the TEI vocabulary is heavily restricted and because it sometimes uses the same elements of the abstract model to serve different text-modeling tasks, it should come as no surprise if some lexicographic expectations get frustrated along the way. A lexicographer may, for instance, object that three distinct lexicographic phenomena — a citation, example and translation equivalent — are lumped together in TEI using just one element (<cit>), albeit with a set of corresponding attributes, but in those cases it is important to keep two things in mind: 1) a serialization is *not* the same thing as a model, and the names of elements in a serialization are, in theory at least, completely arbitrary; and 2) as a general-purpose vocabulary, TEI has to restrict the temptation of overspecialization in order to keep its overall architecture *and* the overall number of elements manageable.

I have established a set of encoding principles to guide me in the process of editing LGSL:

1. mark up the text as precisely, explicitly and consistently as possible, while trying to choose simple over more complex solutions;
2. use the existing TEI tagset and do *not* introduce new TEI elements;
3. when TEI offers choice, use the semantically *most specific* TEI tag available for the given phenomenon, unless this choice would contradict the requirements of precision, explicitness, consistency and simplicity (PECS)
4. modify the TEI schema only if the available TEI mechanism does not satisfy our other requirements (PECS).

Needless to say, editing a dictionary and applying markup to it are activities that do not happen in a temporal vacuum. All decisions are therefore heavily influenced by the editor's basic cost-benefit calculation: how much time and effort can I afford to spend on encoding a particular feature? The calculation is based on one's own research priorities, but also on the projection of a feature's potential use for other researchers or the general public.

5.2 Macrostructure

5.2.1 Entries

TEI offers no less than five different elements for grouping lexical information in a dictionary:

1. `<entry>`: contains a single structured entry in any kind of lexical resource, such as a dictionary or lexicon.
2. `<entryFree>`: contains a single unstructured entry in any kind of lexical resource, such as a dictionary or lexicon.

3. `<superEntry>`: groups a sequence of entries within any kind of lexical resource, such as a dictionary or lexicon, which function as a single unit, for example a set of homographs.
4. `<re>`: (related entry) contains a dictionary entry for a lexical item related to the headword, such as a compound phrase or derived form, embedded inside a larger entry.
5. `<hom>`: groups information relating to one homograph within an entry.

Having five different elements to group lexical items is not a particularly desirable feature, both in terms of the potential for interoperability, and in terms of pedagogical accessibility. While one *could* argue why all these choices are or were necessary, the language and the logic of the *TEI Guidelines* is not very precise here, both at the level of distinguishing structured from unstructured content, and in terms of clearly explaining containing vs. contained elements.

`<entry>` and `<entryFree>` are distinguished as structured vs. unstructured but this is a gross oversimplification. `<entryFree>` is not unstructured but rather allows for more loosely structured content. The content model for `<entry>` is more restrictive: only two elements (`hom` and `sense`) and three classes (`model.entryPart.top`, `model.global` and `model.ptrLike`) are allowed (see Table 4 on p. 152). Apart from those two explicitly mentioned element and the members of the global class (element that can appear at any point within a TEI text), an entry is limited to the so-called top-level entry parts (`cit`, `def`, `dictScrap`, `etym`, `form`, `gramGrp`, `re`, `usg`, and `xr`). `entryFree`, on the other hand, can contain *any* dictionary-specific element,⁶⁴ so the number of elements allowed within an

⁶⁴ i.e. `case`, `colloc`, `def`, `etym`, `form`, `gen`, `gram`, `gramGrp`, `hom`, `hyph`, `iType`, `lang`, `lbl`, `mood`, `number`, `oRef`, `oVar`, `orth`, `pRef`, `pVar`, `per`, `pos`,

entryFree is much larger. All these elements, however, provide structure, so calling entryFree unstructured by default would be a mistake.

Table 4: Differences in content model between <entry> and <entryFree>

entry	entryFree
<pre><content> <alternate minOccurs="1" maxOccurs="unbounded"> <elementRef key="hom"/> <elementRef key="sense"/> <classRef key="model.entryPart.top"/> <classRef key="model.global"/> <classRef key="model.ptrLike"/> </alternate> </content></pre>	<pre><content> <alternate minOccurs="0" maxOccurs="unbounded"> <textNode/> <classRef key="model.gLike"/> <classRef key="model.entryPart"/> <classRef key="model.morphLike"/> <classRef key="model.phrase"/> <classRef key="model.inter"/> <classRef key="model.global"/> </alternate> </content></pre>

The implicit reason why entryFree is mentioned by the *Guidelines* as a counterpoint to entry is because the former, unlike the latter, allows text nodes. So, in principle, and in principle only, one *could* have a valid TEI entry that consisted of <entryFree> and plain text inside. This, however, would not be a particularly useful kind of encoding. Having text nodes in entryFree, however, is important because without text nodes, as we shall see in the next section, it would be impossible to encode punctuation marks according to our encoding principles.

For grouping entries together, TEI has a specific element **superEntry**, which can contain entry but not entryFree. At the same time TEI has an element **hom** (homograph) which “groups information relating to one homograph within an entry” indicating that the functionality of **superEntry** can be replicated within an entry as well. This is another example when too much flexibility on the part of TEI

pron, re, sense, subc, superEntry, syll, tns, usg, and xr.

can lead to confusion and mutually less compatible digital editions of lexicographic content.

That is why in LGSL, macrostructure is encoded only using `<entryFree>`. No grouping is allowed using `<superEntry>`, not only because `superEntry` does not allow `entryFree`, but because, we want to make sure that *all* entries in the dictionary are at the same level in the dictionary hierarchy. This, as we shall see in the next section, applies to homographs as well.

Each entry in the dictionary has a unique id. The unique ids are created automatically by an XSLT script: the id consists of the abbreviation of the author's name, the dictionary title and non-accented lemma:

```
<entryFree xml:id="VSK.SR.грабити">
  <!-- etc. -->
</entryFree>
```

In case the lemma is a reflexive verb or a multiword expression, the spaces in the `@xml:id` are replaced with an underscore:

```
<entryFree xml:id="VSK.SR.грабити_се">
  <!-- etc. -->
</entryFree>
```

Homographs, as we will see in the next section, have their homograph number as part of the unique id.

5.2.2 Homographs

Homographs are lexemes that are orthographically identical to each other but semantically different. In *LSGL*, homographs are presented as separate, numbered entries. This means that each entry has its own lemma, but the number before the lemma indicates that the entry is either preceded by or followed by another homographic entry.

1. **гѣша**, f. 1) der Kropf der Vögel, guttur. — 2) der Kropf der Menschen, struma. — 3) die Kehle, jugulum: ухватио га за гѣшу. cf. [vide] грло[1].
 2. **гѣша**, f. 1) гушава жена, die Kropfige, strumosa. — 2) (ист.) vide гушо.

An example of homographic entries

As the above example shows, E3 treats entries as homographic despite the possible difference in accentuation (гѣша vs. гѣша). In accordance with our encoding principles of precision, explicitness and consistency, homographs are encoded as `entryFree`, like any other entry in the dictionary, with the addition of three attributes: `@type`, `@n` and `@corresp`. The `type` attribute is used to specify that the given entry is indeed, within the concept of a given dictionary, considered a homograph. The `n` attribute contains the number of the homograph in the given dictionary, whereas the `corresp` attribute is used to establish explicit links between homographs using their unique ids:

```
<entryFree xml:id="VSK.SR.гѣша1" type="hom" n="1"
  corresp="#VSK.SR.гѣша2">
  <!-- etc. -->
</entryFree>
<entryFree xml:id="VSK.SR.гѣша2" type="hom" n="2"
  corresp="#VSK.SR.гѣша1">
  <!-- etc. -->
</entryFree>
```

TEI, as we have seen, has a more specific element `hom` to describe homographs, but its use would not contribute to the precise, explicit, consistent and simplest possible encoding that we are trying to produce. With the above encoding, the macrostructure is consistently encoded using one single element (which simplifies,

for instance, the creation of a search engine on top of the XML data), while the homographs are precisely and explicitly singled out with the a set of attributes.

5.3 Microstructure

5.3.1 General remarks

A prototypical *LSGL* entry will look like this:

```
<entryFree>
  <form>
    <!-- word forms -->
  </form>
  <gramGrp>
    <!-- grammar -->
  </gramGrp>
  <sense n="1">
    <!-- etc. -->
  </sense>
  <sense n="2">
    <!-- etc. -->
  </sense>
</entryFree>
```

Despite the fact that we have opted for `entryFree` instead of `entry`, we keep a strict microstructural hierarchy: grouping all form-related elements under `form`; all grammar-related elements under `gramGrp`, and all sense-related elements under `sense`. As we shall see later on, when we delve into the details of each of these groups, we will never allow `orth` outside of `form`, or `def` outside of `sense`, even though `entryFree` would permit such encoding. This is because our goal is to produce consistent structuring of dictionary entries, not only in *LSGL* but across different dictionaries that are and will be published on <http://raskovnik.org> and that for at least two reasons: (1) so that a robust search mechanism can be built on top of dictionaries that, on the surface level, may seem to be structured differently; and (2) in order to facilitate meaningful comparison between different dictionaries.

If we insist on the strict structuring of `<entryFree>`, why didn't we opt for a formally stricter `<entry>` in the first place? The reason is very simple: `entry` cannot contain text nodes, and for good reason. The thinking behind the content model of `entry` is logical: by not allowing any unstructured content, `entry` forces encoders to be explicit about every component of the dictionary entry. The problem with this content model is that it is entirely based — and fully functional — for the so-called *lexical view* of the dictionary: an abstracted, database-like representation of the dictionary content. Legacy dictionaries, however, are not databases: they are historical documents. And our goal should be to encode them as such, i.e. to fully represent their *textual* content. The restricted nature of `entry` makes that job impossible. Consider this simple entry as an example:

Ђѣрѣа, f. Ѣрауеннаме, nomen feminae.

The comma after the lemma and before the part-of-speech is not part of the word-form: it is a conventional typographic separator between the lemma and the part-of-speech. This convention is used in countless dictionaries. Strictly speaking, it is not necessary, one *could* create a perfectly legible and understandable print dictionary without this comma, but Vuk didn't. If we tried to use the comma inside an `<entry>`, our encoding would be invalid:

```
<entry xml:id="VSK.SR.Ђѣрѣа">
  <form type="lemma">
    <orth norm="Ђѣрѣа">Ђѣрѣа</orth>
  </form>, <gramGrp>
    <pos>f.</pos>
  </gramGrp>
  <!-- etc. -->
</entryFree>
```

One would be tempted to encode the comma as a `<pc>` (punctuation character) or as a `<lb1>` considering that comma is a punctuation character whose function in the given case is to serve as a label indicating the separation between the form-related and grammar-related elements. But the content model of `entry` does not allow either: `pc` is a member of two classes: `model.linePart` and `model.segLike`, neither of which is allowed within `<entry>`. `<lb1>`, on the other hand, is a member of `model.entryPart`, but `<entry>` can only take members of the `model.entryPart.top` class, i.e. the top-level entry parts.

Three other options would be possible: to encode the comma within the `<form>`, as a `<dictScrap>` or to not encode it at all. But none of these options would be acceptable:

1. encoding comma within the form (`<form><orth>ḥṽpḥa</orth>, </form><gramGrp><!-- etc. --></gramGrp>`) would be technically valid, but would be semantically problematic because it would imply that the comma is part of the form, which it isn't: the comma doesn't qualify the form in any way — it is external to the form.
2. encoding comma as a `dictScrap` would be technically valid (`<form><orth>ḥṽpḥa</orth></form><dictScrap>, </dictScrap><gramGrp><!-- etc. --></gramGrp>`) because `dictScrap` is a top-level entry element, but this would be a clear case of tag abuse. `dictScrap`, as we shall see later “encloses a part of a dictionary entry in which other phrase-level dictionary elements are freely combined” and is best used as a container for items that don't easily fall under the category of top-level entry elements.

3. removing the comma would spare us some editorial headaches but would falsify the text of the dictionary. In theory, one could omit the comma in XML and produce it in post-processing with a corresponding XSLT stylesheet, but this would still create an encoding that would not reflect the dictionary as a complex assemblage, which is a text, a tool, a model and a cultural artifact at the same time. The encoded dictionary is our master document: we can transform it and style it differently in the online edition, depending on the needs of our user base and our esthetic sensibilities, but the master document in TEI XML should not be a slimmed-down version that has some of its features removed in order to fit the TEI schema.

A fourth option would have been possible: to extend the content model of `entry` to allow text nodes, but this would have gone against the fourth of our Encoding Principles — to modify the schema only if the available mechanisms do not produce precise, explicit, consistent and simple solutions. Using a stricter implementation of `entryFree` (containing top-level entry elements + text nodes for separators between top-level entry elements) satisfies those requirements.

5.3.2 *Forms*

In *LSGL*, forms are typed elements: each form has a `@type` attribute to distinguish it from other forms in the entry. Furthermore, `form` serves as containers for `orth`, which marks up the actual form. This is needed in case a specific word form is further qualified (by a label, or a POS-tag). At the entry level, however, we assume (pace Ide *et al.* 2007) the presence of dependency rules that block the propagation of subordinate features when the privileged feature is overwritten. In other words: the grammatical properties relevant to the entry as a whole are to be found in

entryFree/gramGrp and not in entryFree/form/gramGrp. The grammatical properties of a single word form can be extracted from the sibling gramGrp of the ancestor-or-self of the focus element, unless they are overwritten at a more granular level.

5.3.2.1 *Lemma*

Each entry contains at least one headword (lemma). Lemmas are typographically set apart in the print edition in bold, and in the digital edition by the @type="lemma".

An entry without a lemma would be a contradiction in terms.

```
<entryFree xml:id="VSK.SR.Караманлија">
  <form type="lemma">
    <orth norm="Караманлија">Караманлија</orth>
  </form>,
  <gramGrp>
    <pos>m.</pos>
  </gramGrp>
  <!-- etc. -->
</entryFree>
```

Lemmas in *LSGL* are accented.⁶⁵ In order to simplify the searching and indexing of non-accented lemmas, we have automatically created a normalized version of the lemmas (without the accents) and placed them in the @norm attribute.

5.3.2.2 *Other word forms*

Lemmas are often accompanied by inflected word forms (for instance: genitive for nouns, or first-person singular for verbs), which can be represented in full or using word endings only.

⁶⁵ Traditionally, Serbian has been described as a pitch accent language with lexical stress, although the four pitch accents (long rising, short rising, long falling and short falling) are not represented in the prosodic systems of all the Serbian dialects (Lehiste and Ivić 1986; Smiljanic 2004). Accents are used mostly in dictionaries and grammars, but very rarely in general written discourse. It is not to be expected that users would search *LSGL* or any electronic dictionary using accents.

Inflected word forms

жа́бац, жа́пца, m. Männchen von Frosch, rana mas.

```
<entryFree xml:id="VSK.SR.жабац">
  <form type="lemma">
    <orth norm="жабац">жа́бац</orth>
  </form>, <form type="inflected">
    <orth>жа́пца</orth>
  </form>, <gramGrp>
    <pos>m.</pos>
  </gramGrp>
  <sense>
    <!-- etc. -->
  </sense>
</entryFree>
```

Labeled inflected word forms

If an inflected word form is accompanied by a qualifier, the qualifier is marked up as a <lbl> (label):

жи́р, m. (loc. жи́ру) die Eifeln, glandes: крупно брашно као жир (а ситно као цјена).

```
<entryFree xml:id="VSK.SR.жир">
  <form type="lemma">
    <orth norm="жир">жи́р</orth>
  </form>, <gramGrp>
    <pos>m.</pos>
  </gramGrp> (<form type="inflected">
    <lbl>loc.</lbl>
    <orth>жи́ру</orth>
  </form>) <sense>
    <!-- etc. -->
  </sense>
</entryFree>
```

5.3.2.3 Word endings

Inflected forms that are represented by word endings are marked up with a corresponding @type:

ħorator, a, o, des Schimmels, equi albi.

```
<entryFree xml:id="VSK.SR.ħorator">
  <form type="lemma">
    <orth norm="ħorator">ħòrator</orth>
  </form>, <form type="ending">
    <orth>a</orth>
  </form>, <form type="ending">
    <orth>o</orth>
  </form>,
  <!-- etc. -->
</entryFree>
```

5.3.3 Grammar

Grammatical categories are grouped under `<gramGrp>`. This element has to have at least one element `<pos>` (part of speech).

```
<entryFree xml:id="VSK.SR.говедар">
  <form type="lemma">
    <orth norm="говедар">говèdâr</orth>
  </form>, <form type="inflected">
    <orth>говедápa</orth>
  </form>, <gramGrp>
    <pos>m.</pos>
  </gramGrp>
  <!-- etc. -->
</entryFree>
```

Unlike nouns and verbs, for instance, which typically have explicit grammatical information associated with them (gender for nouns, or aspect for verbs), adjectives throughout *LSGL* are for the most part not labeled as such. For the sake of consistency, we mark them up as adjectives with a `@source` attribute pointing to the bibliographic information of the digital edition. This way we indicate that the explicit marking of the part-of-speech in this instances does not come from any of the print editions:

àбраш.љив, а, о, vide пјегав (н. п. коњ испод
репа).

```
<entryFree xml:id="VSK.SR.абрашљив">
  <form type="lemma">
    <orth norm="абрашљив">àбрашљив</orth>
  </form>, <form type="ending">
    <orth>a</orth>
  </form>, <form type="ending">
    <orth>o</orth>
  </form>, <gramGrp source="#digitalEd">
    <pos>adj.</pos>
  </gramGrp>
  <!-- etc. -->
</entryFree>
```

5.3.4 Etymology

All words of Turkish origin in *LSGL* are marked with an asterisk. To mark them up, we are using elements `<etym>` and `<lang>`. To facilitate searching across different dictionaries, the `@value` attribute is added to `<lang>` with a two-letter (ISO 639-1) code for Turkish (tr).

авдес,* m. die Abwaschung (des Türken) vor dem
Gebete, ablutio: узео авдес; Чисто Турски
авдес узимаше —

```
<entryFree xml:id="VSK.SR.аван">
  <form type="lemma">
    <orth norm="аван">àван</orth>
  </form>,
  <etym>
    <lang value="tr">*</lang>
  </etym>
  <gramGrp>
    <pos>m.</pos>
  </gramGrp>
  <!-- etc. -->
</entryFree>
```

In some entries, etymological labels were only added in E3, inside square brackets.

In those cases <etym> gets an additional @source="#thirdEd".

ђерз[*], [←] m. (у Далм.) као момак, der Burſch,
adolescentulus: Све би ђерзе за два ђерза
дала —

```
<entryFree xml:id="VSK.SR.ђерз">
  <form type="lemma">
    <orth norm="ђерз">ђерз</orth>
  </form>
  <etym source="#thirdEd">
    <lang value="tr">[*]</lang>
  </etym>
  <!-- etc. -->
</entryFree>
```

5.3.5 Usage

Labeling is a traditional lexicographic device for signaling exceptionality. The presence of a usage label implies that “a certain lexical item deviates in some respect from the main bulk of items described in a dictionary and that its use is subject to some kind of restriction” (Svensén 2009: 315). That is why labels are often used in dictionaries to succinctly indicate sociocultural parameters of meaning (Burkhanov 2003), including relations between interlocutors, and their social and cultural roles, attitudes, values and beliefs (Аперсян 1988; Wierzbicka 1992; Marmaridou 2000). In the context of learners’ dictionaries, the use of a labeling system has been seen as particularly important for marking words which are formal or literary, informal or offensive, for instance (Kirkpatrick 1985).

The problem with lexicographic labels in general, however, is they tend to restrict the meaning of the word they are attached to in a definitive and absolute manner (see Тасовац and Башић 2012). Labeling usage with sociolinguistic

indicators such as *SLANG* or *TABOO* is particularly difficult because of the malleable nature of sociolinguistic rules and a wide variety of contexts that need to be taken into consideration. At the same time, dictionary entries — even in the age of electronic lexicography — cannot for practical reasons afford to contain extensive scholarly articles on topics of usage in each labeled entry.

It has been suggested that labels — in their condensed, abbreviated variety — may simply be a remnant of the past:

an artifact of the traditional format of a dictionary: a book in print with a very restricted amount of space. In that sense, (*inf.*) could really be taken as just a shorthand for ‘an informal way of saying’ saving 19 space units in a definition. Since modern technology allows more space to electronic dictionaries, there is an opportunity to work away labels in longer definitions. (Verkuyl *et al.* 2003: 298)

While this may be true in theory (and also applying in general to all abbreviated content in electronic dictionaries, including grammatical information or bibliographic information in citations, for instance), explicit usage notes have one advantage over the blended-in, narrative kind that is especially relevant in the process of encoding legacy dictionaries: they are easier to parse automatically. In *LSGL*, usage notes *tend to* be separated from the rest of the text by parentheses: (*y II. ī.*) = *y Црној њори* [in Montenegro]. This is important also because parsed usage labels can, as I will show in the next chapter, be used as facets for navigating and exploring the dictionary content beyond the single-word look-up paradigm.

In TEI, *<usg>* can be attached at various points in the entry hierarchy in order to signal restrictions of the particular lexical item that it is attached to. In *LSGL*, usage notes are usually either top-level entry elements, in which case they

apply to the entry as a whole; or they restrict only particular senses. In the context of *LSGL*, we have adopted the following rules regarding encoding usage:

1. `@type` is a mandatory attribute on `<usg>`. Without `@type`, `<usg>` would be an underspecified element. Usage labels describe a wide range of linguistic phenomena. Classifying them should be considered a good practice.
2. the values of `@type` that we use are different than the ones suggested by the *Guidelines*. I have adopted a classical lexicographic terminology for classifying usage labels following Hausmann (1989c). See next section for an explanation.
3. I use `<usg>` in a narrower sense than the *Guidelines* suggest. There is too much potential overlap between `<usg>` and some other TEI elements.
4. `@norm` is a mandatory attribute on `<usg>`. Because dictionaries vary in the usage terminology they employ, and because they are not always fully consistent in the use of their own terminology, normalizing `usg` labels or linking them to a controlled vocabulary should be considered best practice.

5.3.5.1 *Types of marking*

Hausmann's (Ibid.) macromodel of diasystemic information contained in usage labels consists of 11 microsystems: a label is understood to be indicating a marked periphery vis-a-vis an unmarked center. If the criterion is time, for instance, the type of marking is called diachronic: its unmarked center is contemporary language, whereas its marked periphery ranges from archaisms to neologisms. The examples of labels that fall under this type of marking would be: *arch*, *dated*, *old* etc. If the criterion is place, the type of marking is diatopic. Its unmarked center is standard

language, its marked periphery: regionalism or dialect word. Examples of diatopic labels would be *AmE*, *Scot.*, *dial etc.* (651) This kind of typology, which explicitly names the criteria, unmarked centers and marked peripheries is superior to the ad hoc values suggested by the *TEI Guidelines*. `<usg type="geo">` is not wrong in itself: after all, the criterion for this type of marking is geographic place or an area. But in the context of dictionaries, in general, and *LSGL*, in particular, geographic labels are given not for the sake of geography, but for the sake of signaling the peripheral nature of a given linguistic realization in relation to an unmarked center.

For this reason, I have adopted Hausmann's terminology in describing usage labels in *LSGL*. In addition to diachronic and diatopic markings, these also include diatextual (based on text type, for instance *poetic*) and diastratic (based on socio-cultural markedness, for instance *urban*). A fifth type (`@type="hint"`) is taken from the Guidelines ("unclassifiable piece of information to guide sense choice") to serve as a catch-all for mostly narrative usage labels. A more detailed classification of these labels could be undertaken in the future.

5.3.5.2 *Diatopic*

Diatopic labels in *LSGL* can occur as top-level entry elements or can be attached to particular senses. They are often abbreviated, appear in parentheses, and contain a reference to a toponym: (*у Ц. љ.*) [in Montenegro], (*у Дубр.*) [in Dubrovnik], (*у Срему*) [in Srem] etc. We are using `<usg type="diatopic">` as a container for a `<rs>` (referring string) which links the occurrence of a place name in the text to a corresponding authority record. The geolocated authority records are kept in a database. To establish an explicit link, it is sufficient to encode the id of the authority record. For instance:

амáнат[*], м. (у Ц. г.) 1) vide аманет. — 2) Црногорци се куну и заклињу аманатом: *аманата* ми Божјега! Кад другога кога заклињу онда се у говору често *та* изостави, па се говори н. п. *амана ти* (мјесто *аманата ти*).

```
<entryFree xml:id="VSK.SR.аманат">
  <form type="lemma">
    <orth norm="аманат">амáнат</orth>
  </form>
  <etym source="#thirdEd">
    <lang value="tr">[*]</lang>
  </etym>,
  <gramGrp>
    <pos>м.</pos>
  </gramGrp>
  <usg type="diatopic">(y <rs key="930" type="geo">Ц. г.</rs></usg>
  <sense n="1" xml:id="VSK.SR.аманат.1">
    <!-- etc. -->
  </sense>
  <sense n="2" xml:id="VSK.SR.аманат.2">
    <!-- etc. -->
  </sense>
</entryFree>
```

This way there is no need for normalizing the toponym locally, since all the forms (abbreviated, semi-abbreviated, full) that occur in the text are linked to the same authority record. All of the following examples appear in the text:

```
<usg type="diatopic">(y <rs key="930" type="geo">Ц. г.</rs></usg>

<usg type="diatopic">(y <rs key="930" type="geo">Црн. гори</rs></usg>

<usg type="diatopic">(y <rs key="930" type="geo">Црној гори</rs></usg>
```

5.3.5.3 Diatextual

Diatextual markings, in the context of *LSGL*, are used to indicate the use of a given word in a particular folk literary genre: у *ѿјесмама* [in poems], у *ѿповијецѿ* [in a novella] etc.

живодерац, дѣрца, m. der lebendig geſchunden wird, qui excoſiatur vivus: „ја сам јарац *живодерац*, жив дрт не одрт, жив печен не испечен“ и т. д. (у приповијецѿ).

`<usg norm="prose" type="diatextual">у приповијецѿ</usg>`

The labeling is not consistent throughout the dictionary: the literary genres are sometimes in the singular, sometimes in the plural. Sometimes, the label appears on its own in brackets (like in the example above), at other times, it is part of a larger explanation, when for instance a “poetic” word is contrasted with its “prosaic” counterpart (often with the formula: “in poems instead of X”). For all these reasons, it was necessary to normalize the values of diatextual labels to ensure consistency.

@norm	Examples from the text
literary ⁶⁶	<code><usg norm="literary" type="diatextual">(у књижевника)</usg></code> <code><usg type="diatextual" norm="literary">(код књижевника)</usg></code> <code><usg type="diatextual" norm="literary" rend="italic">(понајвише код књижевника)</usg></code>

⁶⁶ *Literary* here is to be understood differently from the other genres. In *LSGL*, all the other genres are referring to folk literary expression, i.e. folk poetry, folk prose etc., which remained mainly oral until the 19th century. *Literary* proper refers to established published literature.

poetry	<usg norm="poetry" type="diatextual">у пјесмама</usg> <usg type="diatextual" norm="poetry">у пјесмама кашто не значи ништа него се додаје само да се испуни врста</usg> <usg norm="poetry" type="diatextual">у пјесми</usg> <usg norm="poetry" type="diatextual">у пјесми мјесто <ref target="#VSK.SR.бјегати" rend="italic"><rs type="lemma">бјегати</rs></ref> да би се врста испунила</usg> <usg norm="poetry" type="diatextual">у припјеву</usg> <usg norm="poetry" type="diatextual">У једној пјесми може бити само стиха ради зовесеи<ref target="#VSK.SR.црквено_коло" rend="italic"><rs type="lemma">коло црквено</rs></ref></usg> <usg norm="poetry" type="diatextual">понајвише у пјесмама</usg>
prose	<usg norm="prose" type="diatextual">у приповијечи</usg> <usg norm="prose" type="diatextual">у приповијеткама</usg> <usg type="diatextual" norm="prose">(по приповијеткама народним)</usg> <usg type="diatextual" norm="prose" xml:lang="de">(in der Anekdote)</usg> <usg norm="prose" type="diatextual">у приповијечи мјесто <ref target="#VSK.SR.крста1<rs type="lemma">крсти</rs></ref></usg>
proverb	<usg norm="proverb" type="diatextual">у овој пословици</usg> <usg norm="proverb" type="diatextual">у пословици</usg> <usg type="diatextual" norm="proverb">у пословици мјесто <ref target="#VSK.SR.давалац" rend="italic"><rs type="lemma">давалац</rs></ref></usg> <usg type="diatextual" norm="proverb">У овој пословици говори се овако и по осталијем крајевима према <ref target="#VSK.SR.рило" rend="italic"><rs type="lemma">рило</rs></ref></usg> <usg type="diatextual" norm="proverb">И од тога је остала ова пословица</usg>
riddle	<usg norm="riddle" type="diatextual">у загонци</usg> <usg norm="riddle" type="diatextual">нијесам чуо до у овој загонци</usg> <usg norm="riddle" type="diatextual">у овој загонци</usg> <usg norm="riddle" type="diatextual">само у овој загонци</usg>
spell	<usg type="diatextual" norm="spell">у овоме врачању</usg> <usg type="diatextual" norm="spell">У врачању</usg>
toast	<usg type="diatextual" norm="toast">(у здравици)</usg>

With normalized values of diatextual markings, it becomes trivial to retrieve, for instance, all entries containing a particular type of marking or to limit one's search only on those entries.

5.3.5.4 *Diaphasic*

In the Preface to E1, Vuk describes this particular type of marking like this:

Које се ријечи данас не говоре у говору, него се само чују у пјесмама, оне сам забележио са ст. (стајаћа, poetisch), као н. п. *чеґо, људа, чарни, сѣаґо, лаґо, лељо*, и т. д.

[The words that are nowadays not heard speech, but only in poems/songs I marked as ст. (стајаћа, poetic), for instance: *чеґо, људа, чарни, сѣаґо, лаґо, лељо*, etc.]

Стајаћи (literally “standing”) is not a common term in Serbian linguistic or lexicographic terminology. The entry for *сѣајаћу* in the dictionary itself has only one sense. It consists of a hint („н. п. хаљина” [e.g. dress]) and translation equivalents (“feierlich, sonntäglich, solemn”). In other words, in the dictionary itself Vuk translates *сѣајаћу* with *ceremonial* or *solemn*, and not *poetic* as he did in the introduction. While entries marked as *сѣајаћу* are undoubtedly connected to examples from folk poetry, Vuk is here stressing two different things: (a) the fact that these words are no longer used in speech („које се ријечи данас не говоре у говору”), and (b) based on the label that he chose, that they “feel” ceremonial or solemn.

The label *сѣајаћу* exemplifies the difficulty of formally marking up usage labels in dictionaries when the labels themselves are ambiguous. The fact that some words are no longer heard in everyday speech would indicate a diamedial marking, whose unmarked center is neutral, and whose marked peripheries are spoken and

written language. But the “ceremonial” aspect is more aligned with diaphasic markings of the level of formality. For practical reasons and in order to distinguish this label from diatextual labels in *LSGL*, we have opted to encode it as `<usg type="diaphasic" norm="elevated">(ст.)</usg>`. It should be mentioned, however, that `@norm="elevated"` should be considered a local normalization and that any attempt to use this normalization across different dictionaries would necessarily require a detailed metalexicographic comparison between the dictionaries.

жарко сунце, н. (ст.) heißen Sonne, sol calidus:
Жарко га је огријало сунце —

```
<entryFree xml:id="VSK.SR.жарко_сунце">
  <form type="lemma">
    <orth norm="жарко сунце">жâрко сѹнце</orth>
  </form>,
  <gramGrp>
    <pos>n.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.жарко_сунце.1">
    <usg type="diaphasic" norm="elevated">(ст.)</usg>
    <cit type="translation">
      <quote xml:lang="de">heiße Sonne</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">sol calidus</quote>
    </cit>:
    <cit type="example">
      <quote>Жарко га је огријало сунце</quote>
    </cit>.
  </sense>
</entryFree>
```

5.3.5.5 Diastratic

In *LSGL*, diastratic markings are used to label urban usage (*у варошима, осодиуо уо варошима, у осѣалим варошима, уо груѣјем онуда варошима* etc.)

**шми́зла, f. (у војв. по варошима) Schemisette:
И одвели шлинговану *шмизлу* —**

```
<entryFree xml:id="VSK.SR.шмизла">
  <form type="lemma">
    <orth norm="шмизла">шми́зла</orth>
  </form>,
  <gramGrp>
    <pos>f.</pos>
  </gramGrp>
  <usg type="diatopic">(у <rs key="164" type="geo">војв.</rs></usg>
    usg>
  <usg type="diastratic" norm="urban">по варошима)</usg>
  <sense xml:id="VSK.SR.шмизла.1">
    <cit type="translation">
      <quote xml:lang="de">Schemisette</quote>
    </cit>:
    <cit type="example">
      <quote xml:lang="sr">И одвели шлинговану <hi
        rend="italic">шмизлу</hi>
      </quote>
    </cit>.
  </sense>
</entryFree>
```

5.3.5.6 Hint

All other grammatical, semantic or contextual usage notes are encoded as <usg type="hint"/>.

**ђурђе́вски, kâ, kô, н. п. мјесец, киша, Georgi-,
S. Georgii.**

```
<entryFree xml:id="VSK.SR.ђурђе́вски">
  <form type="lemma">
    <orth norm="ђурђе́вски">ђурђе́вски</orth>
  </form>,
  <form type="ending">
    <orth>kâ</orth>
  </form>,
  <form type="ending">
    <orth>kô</orth>
  </form>,
  <gramGrp source="#digitalEd">
    <pos>adj.</pos>
```

```

</gramGrp>
<sense xml:id="VSK.SR.ђупђевски.1">
  <usg type="hint">н. п. мјесец, киша</usg>,
  <cit type="translation">
    <quote xml:lang="de">Georgi-</quote>
  </cit>,
  <cit type="translation">
    <quote xml:lang="la">S. Georgii</quote>
  </cit>.
</sense>
</entryFree>

```

5.3.5.7 Hint in German

Hints can also occur in German translations. They are marked the same way as hints in Serbian, with the exception of a mandatory language attribute.

жѣстити се, жѣстѣм се, v. r. impf. entbrennen
з. В. vor Zorn, excandescio ira.

```

<entryFree xml:id="VSK.SR.жестити_се">
  <form type="lemma">
    <orth norm="жестити се">жѣстити се</orth>
  </form>,
  <form type="inflected">
    <orth>жѣстѣм се</orth>
  </form>,
  <gramGrp>
    <pos>v. r. impf.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.жестити_се.1">
    <cit type="translation">
      <quote xml:lang="de">entbrennen</quote> <usg type="hint"
        xml:lang="de">z. B. vor Zorn </usg>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">excandescio ira</quote>
    </cit>.
  </sense>
</entryFree>

```

5.3.6 Senses

Even though <sense> is not strictly required by <entryFree>, we have made it mandatory for all dictionary entries. This not only reflects the underlying

semasiological model of the dictionary, but actually leads to more consistency across various dictionaries. Senses are automatically assigned unique xml:ids based on the entry id and the sense number.

ајлукчија,* м. војник који војује за плату, ein besoldeter Krieger, miles mercenarius.

```
<entryFree xml:id="VSK.SR.ајлукчија">
  <form type="lemma">
    <orth norm="ајлукчија">ајлукчија</orth>
  </form>,
  <etym>
    <lang value="tr">*</lang>
  </etym>
  <gramGrp>
    <pos>м.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.ајлукчија.1">
    <def>војник који војује за плату</def>, <cit
      type="translation">
      <quote xml:lang="de">ein besoldeter Krieger</quote>
    </cit>, <cit type="translation">
      <quote xml:lang="la">miles mercenarius</quote>
    </cit>. </sense>
</entryFree>
```

Senses may contain the following types of lexicographic information:

- definition (see Section 5.3.6.1, pp. 175-175)
- translation equivalents (see Section 5.3.6.2, pp. 175-179)
- examples (see Section 5.3.6.3, pp. 179-180)
- additional information (see Section 5.3.6.4, pp. 180-181)
- usage information (see Section 5.3.5, pp. 163-173)
- cross-references (see Section 5.4, pp. 186-208)

5.3.6.1 Definition

We use <def> to mark up lexicographic definitions. No `xml:lang` attribute is required because of the inheritance rules.

ђаволица, f. жена, или дјевојка, која ђаволи,
die Teufelin (muthwilliges Frauenzimmer), femina
petulantior, cf. враголица: Већ девојке *ђа-
волице* —

```
<entryFree xml:id="VSK.SR.ђаволица">
  <form type="lemma">
    <orth norm="ђаволица">ђаволица</orth>
  </form>,
  <gramGrp>
    <pos>f.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.ђаволица.1">
    <def>жена, или дјевојка, која ђаволи</def>,
    <!-- etc. -->
  </sense>
</entryFree>
```

5.3.6.2 Translations

For each translation into German and Latin, we use the container element <cit type="translation"> and, within it, a <quote> with a corresponding `@xml:lang`.

запатити, тѣм, v. рѣ. што, н. п. свиња, коза,
оваца, kommen zu etwaz, aufbringen, acquirō,
comparo. cf. [заврѣи 2,] заметнути 4.

```
<entryFree xml:id="VSK.SR.запатити">
  <form type="lemma">
    <orth norm="запатити">запатити</orth>
  </form>,
  <form type="ending">
    <orth>тѣм</orth>
  </form>,
  <gramGrp>
```

```

    <pos>v. pf.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.запатити.1">
    <usg type="hint"> што, н. п. свиња, коза, оваца</usg>,
    <cit type="translation">
      <quote xml:lang="de">kommen zu etwas</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="de">aufbringen</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">acquirō</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">comparō</quote>
    </cit>.
    <!-- etc. -->
  </sense>
</entryFree>

```

Translations into German and Latin sometimes contain foreign words. These foreign words are marked up with a general-purpose container `<seg>` (segment) inside `<quote>` and given a corresponding `@xml:lang`.

Greek in Latin

The Greek neuter article $\tau\omicron$ is used in combination with a Latin infinitive for the translation of Serbian verbal nouns. For instance:

ерендисање, н. das Hobeln, τὸ ῥυψινᾶρε. [cf. рендеисање, рендисање].

```

<entryFree xml:id="VSK.SR.ерендисање">
  <form type="lemma">
    <orth norm="ерендисање">ерендисање</orth>
  </form>,
  <gramGrp>
    <pos>n.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.ерендисање.1">
    <cit type="translation">
      <quote xml:lang="de">das Hobeln</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la"><seg xml:lang="el">τὸ</seg></quote>
    </cit>
  </sense>
</entryFree>

```



```

        runcinare
      </quote>
    </cit>.
    <!-- etc. -->
  </sense>
</entryFree>

```

Greek and Serbian in Latin

Greek definite articles in the genitive case (τοῦ, τῆς, τοῦ) are used to indicate possessive relations in some Latin translations. For instance: *Ђурђевина* is defined in Serbian as „оно што је Ђурђево (било, т. ј. Ђурђа Смедереваца” [that which belongs to Đurađ (used to belong, i.e. to Đurađ Smederevac)]⁶⁷. In German, this is rendered as “das Gebiet des Ђурађ Смедеревац” and in Latin as “imperium тоῦ Ђурађ Смедеревац”.

Ђурђевина, f. (ст.) оно што је Ђурђево (било, т. ј. Ђурђа Смедереваца), das Gebiet des Ђурађ Смедеревац, imperium тоῦ Ђурађ Смедеревац: Који ми је све моје узео: Ђурђевину и Смедеревину —

```

<entryFree xml:id="VSK.SR.Ђурђевина">
  <form type="lemma">
    <orth norm="Ђурђевина">Ђурђевина</orth>
  </form>,
  <!-- etc. -->
  <sense xml:id="VSK.SR.Ђурђевина.1">
    <!-- etc. -->
    <cit type="translation">
      <quote xml:lang="de">das Gebiet des <seg xml:lang="sr">Ђурађ
        Смедеревац</seg>
      </quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">imperium <seg xml:lang="el">τοῦ</seg>
        <seg xml:lang="sr">Ђурађ Смедеревац</seg>
      </quote>
    </cit>
  </sense>
</entryFree>

```

⁶⁷ Đurađ Branković, also known as Smederevac (= of Smederevo), was the Serbian despot from 1427 to 1456 (see Спремић 1994).

```

    <!-- etc. -->
  </sense>
</entryFree>

```

Serbian in German

In the previous example, we already saw that Serbian words are sometimes part of German translations. If the Serbian segment can also be identified as an entry in the dictionary, a cross-reference is established.

жуњин, а, о, der жуња gehörig, pici.

```

<entryFree xml:id="VSK.SR.жуњин">
  <form type="lemma">
    <orth norm="жуњин">жуњин</orth>
  </form>
  <!-- etc. -->
  <sense xml:id="VSK.SR.жуњин.1">
    <cit type="translation">
      <quote xml:lang="de">der <seg xml:lang="sr">
        <ref target="#VSK.SR.жуња">
          <rs type="lemma">жуња</rs>
        </ref>
      </seg> gehörig</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">pici</quote>
    </cit>.
  </sense>
</entryFree>

```

On cross-references in general, see Section 5.4, pp. 186–208.

German explanations in Latin

If a Latin translation contains additional explanation in German, this, too, is treated like a segment with a corresponding @xml:lang.

**арка,* f. der Rücken, dorsum (im Kleide). cf. [vide]
леђа.**

```

<entryFree xml:id="VSK.SR.арка">
  <form type="lemma">
    <orth norm="арка">арка</orth>
  </form>

```

```

<!-- etc. -->
<sense xml:id="VSK.SR.арка.1">
  <cit type="translation">
    <quote xml:lang="de">der Rücken</quote>
  </cit>,
  <cit type="translation">
    <quote xml:lang="la">dorsum <seg xml:lang="de">(im Kleide)</seg>
    </quote>
  </cit>.
<!-- etc. -->
</sense>
</entryFree>

```

5.3.6.3 Examples

Examples are marked up with <cit @type="example"> and an embedded <quote>.

We do not use the xml:lang attribute on <quote> because of the inheritance rules.

алај,* алаја, m. (gen. pl. алаја) ein Trupp Soldaten in Parade, acies instructa, pompa militaris: Туд' ће поћи војска на *алаје* — Нареди се *алај* за *алајем* —

```

<entryFree xml:id="VSK.SR.алај">
  <form type="lemma">
    <orth norm="алај">алај</orth>
  </form>
  <!-- etc. -->
  <sense xml:id="VSK.SR.алај.1">
    <cit type="translation">
      <quote xml:lang="de">ein Trupp Soldaten in Parade</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">acies instructa</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">pompa militaris</quote>
    </cit>:
    <cit type="example">
      <quote>Туд' ће поћи војска на <hi rend="italic">алаје</hi></quote>
    </cit> – <cit type="example">
      <quote>Нареди се <hi rend="italic">алај</hi> за <hi rend="italic">алајем.</hi></quote>
    </cit>
  </sense>
</entryFree>

```

```
</sense>
</entryFree>
```

5.3.6.4 Additional content

In addition to definitions, translations and examples, entries may occasionally contain additional information of encyclopedic rather than purely lexicographic nature. In *LSGL* these are often Vuk's ethnographic notes describing customs, superstitions, various anecdotes etc. This type of narrative is encoded using `<dictScrap>`.

**авгута̀р, августáра, м. сух ри́бљи [суха ри́бља]
мријест, који [која] се (понајвише у Скадру)
вади из рибе скакавице, Art Pökelrogen, ova
piscium condita. Авгута̀р се једе и уз часни
пост сваки дан, као и хајвар.**

```
<entryFree xml:id="VSK.SR.авгута̀р">
  <form type="lemma">
    <orth norm="авгута̀р">авгута̀р</orth>
  </form>
  <!-- etc. -->
  <sense xml:id="VSK.SR.авгута̀р.1">
    <def>сух ри́бљи [суха ри́бља] мријест, који [која] се (понајвише
      у Скадру) вади из рибе скакавице</def>,
    <cit type="translation">
      <quote xml:lang="de">Art Pökelrogen</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">ova piscium condita</quote>
    </cit>.
    <dictScrap>Авгута̀р се једе и уз часни пост сваки дан, као и
      хајвар.</dictScrap>
  </sense>
</entryFree>
```

Some entries may not have a lexicographic definitions at all but a short narrative which illustrates and explains the lemma. Since we reserve the use of `<def>` for definitions proper, narrative segments are encoded as `<dictScrap>` even when they replace a more traditional lexicographic definition:

жва́калица, f. приповиједа се да су некаки Турци, зликовци, пошто их људи почасте искали још да им плате и *жвакалицу*, т. ј. што су се трудили и жватали. И сад кад ко којешта иште много, рече му се: још да ти платим и жвакалицу. cf. озубина, одрина [2, одриња].

```
<entryFree xml:id="VSK.SR.жвакалица">
  <form type="lemma">
    <orth norm="жвакалица">жва́калица</orth>
  </form>
  <!-- etc. -->
  <sense xml:id="VSK.SR.жвакалица.1">
    <dictScrap>приповиједа се да су некаки Турци, зликовци, пошто
    их људи почасте искали још да им плате и <hi
    rend="italic">жвакалицу</hi>, т. ј. што су се трудили и
    жватали. И сад кад ко којешта иште много, рече му се: још да
    ти платим и жвакалицу.</dictScrap>
    <xr type="cf">
      <!-- etc. -->
    </xr>.
  </sense>
</entryFree>
```

5.3.7 Names and named entities

LSGL is a rich source of information about names and named entities. They are encoded at two different points in the microstructure:

1. entries that are themselves toponymic or anthroponymic are marked as such for easy reference
2. occurrences of person and place names in the definitions and examples are also marked up

5.3.7.1 Toponyms and anthroponyms as entries

Entries describing toponyms and anthroponyms are marked up with empty elements `<lbl source="#digitalEd" norm="toponym"/>` and, correspondingly,

<lbl source="#digitalEd" norm="anthroponym"/>. The source attribute unmistakably identifies these labels as originating in the digital edition of the dictionary. If the label occurs at the entry level, it applies to the whole entry. If it occurs at the sense level, it only applies to the given sense. The following example shows an entry that has both different parts of speech and different onomastic labels in each of the senses: in the first sense, *Плавша* is a masculine name of a person; in the second sense, it is a feminine name of a mountain.

Плѣвша, 1) м. Маннынаме, номен viri. — 2) ф. планина у Турској Хрватској: Он ме јесте један пут гонио, | Преко Плавше високе планине —

```
<entryFree xml:id="VSK.SR.Плавша">
  <form type="lemma">
    <orth norm="Плавша">Плѣвша</orth>
  </form>
  <sense n="1" xml:id="VSK.SR.Плавша.1">
    <lbl type="senseNum">1</lbl>
    <gramGrp>
      <pos>m.</pos>
    </gramGrp>
    <lbl source="#digitalEd" norm="anthroponym"/>
    <cit type="translation">
      <quote xml:lang="de">Mannsname</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">nomen viri</quote>
    </cit>.
  </sense>
  <sense n="2" xml:id="VSK.SR.Плавша.2">
    <lbl type="senseNum">2</lbl>
    <gramGrp>
      <pos>f.</pos>
    </gramGrp>
    <lbl source="#digitalEd" norm="toponym"/>
    <def>планина у <rs type="geo" subtype="place">Турској
      Хрватској</rs>
    </def>:
    <cit type="example">
      <quote xml:lang="sr">Он ме јесте један пут гонио, | Преко <hi
```

```

rend="italic">Плавше</hi> високе планине</quote>
</cit>.
</sense>
</entryFree>

```

5.3.7.2 Places

We have already encountered named places in the discussion of diatopic usage labels (Section 5.3.5.2, pp. 166-167). They are marked as <rs> (referring strings) which, when available, point to entries in our authority records. The same principle applies when toponyms appear elsewhere in the entry.

Бањска̂, f. adj. варош у Косову, eine Stadt in
Косово, nomen urbis: Од малене *Бањске*
крај Косова —

```

<entryFree xml:id="VSK.SR.Бањска">
  <form type="lemma">
    <orth norm="Бањска">Бањска̂</orth>
  </form>,
  <gramGrp>
    <pos>f. adj.</pos>
  </gramGrp>
  <lbl source="#digitalEd" norm="toponym"/>
  <sense xml:id="VSK.SR.Бањска.1">
    <def>варош у <rs type="geo" key="420">Косову</rs>
    </def>,
    <cit type="translation">
      <quote xml:lang="de">eine Stadt in <rs type="geo" key="420"
        xml:lang="sr">Косово</rs>
      </quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">nomen urbis</quote>
    </cit>:
    <cit type="example">
      <quote xml:lang="sr">Од малене <rs type="geo" rend="italic"
        key="2978">Бањске </rs>крај <rs type="geo" key="420">Косова</
        rs>
      </quote>
    </cit>.
  </sense>
</entryFree>

```

5.3.7.3 Persons

Persons, both historical and fictitious, are marked up using the same construct we've seen in the previous section, only with a different attribute value: `<rs type="person"/>`. Historical persons are linked to the Raskovnik Named Entities Authority Records (NEAR).

**бупило, н. Приповиједа се и пјева да је цару
Лазару, кад је први пут дошао напомот Раваници,
зазreo коњ од ње и збацио га са себе и да се је то мјесто од онда прозвало
Царево бупило: Ту се зове Царево бупило —**

```
<entryFree xml:id="VSK.SR.бупило">
  <form type="lemma">
    <orth norm="бупило">бупило</orth>
  </form>,
  <gramGrp>
    <pos>n.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.бупило.1">
    <dictScrap>Приповиједа се и пјева да је <rs type="person"
      key="5122">цару Лазару</rs>, кад је први пут дошао напомот <rs
      type="geo" key="706">Раваници</rs>, зазreo коњ од ње и збацио
      га са себе и да се је то мјесто од онда прозвало <hi
      rend="italic">Царево бупило</hi>:
    </dictScrap>:
    <cit type="example">
      <quote xml:lang="sr">Ту се зове Царево бупило.</quote>
    </cit>
  </sense>
</entryFree>
```

5.3.7.4 Vuk's self-reference

References to persons include Vuk's self-references, i.e. those instances in the dictionary text in which Vuk either speaks in the first person singular or mentions situations or events that he was involved in personally. A classic example would be:

the polysemous entry *вук* [wolf], which, in the second sense, defines Vuk as a personal name and Vuk recounts the story of why certain babies, including himself, are given this name:

вѹк, m. 1) der Wolf, lupus. [cf. звијерац, 2 ка-
мењак, курјак]. — 2) [*Вук*] Mannsname, no-
men viri. Кад се каквој жени не даду дјеца
онда надјене дјетету име Вук, јер мисле да
им дјецу вјештице једу, а на *вука* да не ће
смјети ударити (за то су и мени овако име
надјели). — 3) (у Ц. г.) у пушке оно

гдје стоји кремен. vide [чeљycka 1] чeљy-
сти, чeљycke.

```
<entryFree xml:id="VSK.SR.вук">
  <form type="lemma">
    <orth norm="вук">вѹк</orth>
  </form>,
  <gramGrp>
    <pos>m.</pos>
  </gramGrp>
  <sense n="1" xml:id="VSK.SR.вук.1">
    <!-- etc. -->
  </sense>
  <sense n="2" xml:id="VSK.SR.вук.2">
    <lbl source="#digitalEd" norm="anthroponym"/>
    <lbl type="senseNum">2</lbl>
    [<form type="lemma" source="#thirdEd" rend="italic">
      <orth norm="Вук">Вук</orth>
    </form>]
    <cit type="translation">
      <quote xml:lang="de">Mannsname</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">nomen viri</quote>
    </cit>.
    <dictScrap>Кад се каквој жени не даду дјеца онда надјене
      дјетету име Вук, јер мисле да им дјецу вјештице једу, а на
      <hi rend="italic">вука</hi> да не ће смјети ударити (за то су
      и <rs type="person" key="4966">мени</rs> овако име надјели)
```

```

    </dictScrap>.
  </sense>
  <sense n="3" xml:id="VSK.SR.Byk.3">
    <!-- etc. -->
  </sense>
</entryFree>

```

The contents of <dictScrap> read as follows:

When a woman can't have children, she names the child Vuk [Wolf] because people think that it is witches who eat their children, and that they will be afraid to attack a *wolf* (which is why this name was given to me too).

5.4 Mediostructure

The mediostructure is a cross-referencing system used to establish relations among different knowledge-bearing dictionary components (“wörterbuchinterne Mediostruktur”) or between dictionaries and other types of external sources (“intertextuelle Mediostruktur”) (Wiegand 1994/2000: 1164). In practical terms, the mediostructure assists the dictionary user in navigating the dictionary and extracting information from different locations, while establishing and strengthening the cohesion of the dictionary content by optimally organizing its access structure (see Gouws and Prinsloo 1998).

LSGL employs only the dictionary-internal mediostructure. Even though Vuk cites at length examples from folk poetry and, sometimes, prose, those citations are never referenced.⁶⁸ The dictionary-internal mediostructure in *LSGL* can be further divided into: explicit cross-references (those labeled as such, for instance *cf.*

⁶⁸ An automatic recognition and attribution of cited materials against a corpus of Serbian folk literature would be a worthwhile exercise now that the full electronic edition of the *Lexicon* has been released.

necessarily contain lexical material from different dialects, but for that to happen, one of such dialects would need to be accepted as the literary standard:

Ако кад Срби у писању књига приме једно макар које од својих нарјечија, онда у рјечницима не ће ни требати ријечи по свакоме нарјечију писати за себе, него само по ономе којијем се књиге успишу, макар назначивши код сваке ријечи како се говори по другијем нарјечијама, н. п. дјед (јужно ђед, источно дед, западно дид)... Али за сад док је још главна брига и потреба да познамо свој народни језик по свијем крајевима, мислим да је ваљало писати све овако као што сам у овој књизи писао. (Ibid.: unpaginated)

[If Serbs ever adopt one of their dialects for writing books, then there will be no need to write separately the words in every dialect in a dictionary, but only in the dialect which is used for writing, possibly indicating next to each word how it is spoken in other dialects, for instance *đed* (southern *ġeg*, Eastern *geg*, Western *gug*)... But for the time being, while our main concern and need is to learn about the vernacular in all the regions, I think it was necessary to write everything the way I did in this book.]

Vuk's use of *prescriptive cross-references* is an attempt at provisionally reconciling two different lexicographic motivations: getting to know the vernacular language in all the regions [„потреба да познамо свој народни језик по свијем крајевима”] and working toward establishing a dialect “most fitting” to become the basis of a literary standard.

Synonymy (along with antonymy) is usually described as a non-hierarchical, symmetric relation between words (Murphy 2003), but establishing differences between near-synonyms is a complex matter (for a discussion of various dimensions of variation, see for instance Edmonds and Hirst 2002). In *LSGL*, *cf.* and *vide* do not establish denotative, stylistic, emotive or collocational variations

between synonyms, but rather establish preferences based on the standardizing potential of a word — from Vuk’s subjective point of view.

We encode all cross-references using `<xr>` (cross reference) as a container which groups together information related to a cross reference:

1. explicit label (`<lbl>`) indicating the type of reference;
2. reference proper (`<ref>`) with a `target` attribute pointing to the `xml:id` of the target object (entry or specific sense); and consisting of one or more typed `<rs>` (reference string) elements, containing various strings that make up a reference (homograph number, if applicable; lemma proper, and sense number, if applicable)

Encoding cross-references in this manner ascertains both the human-legible representation of the way cross-references appear in the print edition, and the explicit machine-actionable representation of those links in the online-edition. The following sections will make the need for such a distinction clearer.

5.4.1.1 Cf. (*confer*)

**же̑ѣти, же̑ѣм, v. impf. (јуж.) wünsch. supio.
cf. жудјети.**

```
<entryFree xml:id="VSK.SR.же̑ѣти">
  <form type="lemma">
    <orth norm="же̑ѣти">же̑ѣти</orth>
  </form>,
  <form type="inflected">
    <orth>же̑ѣм</orth>
  </form>,
  <gramGrp>
    <pos>v. impf.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.же̑ѣти.1">
    <usg type="diatopic">(<rs type="geo">јуж.</rs></usg>
    <cit type="translation">
      <quote xml:lang="de">wünschen</quote>
```

```

</cit>,
<cit type="translation">
  <quote xml:lang="la">cupio</quote>
</cit>.
<xr type="cf">
  <lbl>cf.</lbl>
  <ref target="#VSK.SR.жудјети">
    <rs type="lemma">жудјети</rs>
  </ref>
</xr>.
</sense>
</entryFree>

```

5.4.1.2 Cf. added in E3

When a cross-reference has been added in E3, it is placed in square brackets. In the digital edition, this type of cross-reference is given an additional attribute (@source="#thirdEd") in order to signal its provenance.

ждрибѣнце (ждрѣбѣнце), цета, п. дѣм. в. ждрибе
[cf. ждрибешце].

```

<entryFree xml:id="VSK.SR.ждрибенце">
  <form type="lemma">
    <orth norm="ждрибенце">ждрибѣнце</orth>
  </form>
  (<form type="lemma">
    <orth norm="ждрибенце">ждрѣбѣнце</orth>
  </form>),
  <form type="ending">
    <orth>цета</orth>
  </form>,
  <gramGrp>
    <pos>n.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.ждрибенце.1">
    <xr type="dim">
      <lbl>dim. <seg xml:lang="de">v.</seg></lbl>
      <ref target="#VSK.SR.ждрибе">
        <rs type="lemma">ждрибе</rs>
      </ref>
    </xr>
    [<xr source="#thirdEd" type="cf">
      <lbl>cf.</lbl>
      <ref target="#VSK.SR.ждрибешце">
        <rs type="lemma">ждрибешце</rs>

```

```

    </ref>
  </xr>].
</sense>
</entryFree>

```

5.4.1.3 Cf. with a specific sense

Cross-references that point to a specific sense indicate the sense number *after* the target lemma in the print edition. For instance, *зайаїиїиїи* is cross-referenced with a fourth sense of *заметнуїиїи*:

зàпатити, тїм, в. рї. што, н. ц. свиња, коза,
овада, łomien zu etwas, aufbringen, acquiro,
comparo. cf. [заврїи 2,] заметнуїи 4.

This machine-actionable link, as in all the previous simpler cases, is encoded in the target attribute of <xr>, whereas the <ref> element gets two <rs> elements: one for the lemma, and one for the sense number:

```

<entryFree xml:id="VSK.SR.зàпатити">
  <form type="lemma">
    <orth norm="зàпатити">зàпатити</orth>
  </form>
  <!-- etc. -->
  <sense xml:id="VSK.SR.зàпатити.1">
    <!-- etc. -->
    <xr type="cf">
      <lbl>cf.</lbl>
      [<ref target="#VSK.SR.заврїи.2" source="#thirdEd">
        <rs type="lemma">заврїи</rs>
        <rs type="senseNum">2</rs>
      </ref>],
      <ref target="#VSK.SR.заметнуїи">
        <rs type="lemma">заметнуїи</rs>
        <rs type="senseNum">4</rs>
      </ref>
    </xr>.
  </sense>
</entryFree>

```

5.4.1.4 Cf. with a specific sense added in E3

As we already saw in Section 5.4.1.2 above, cross-reference that have been added in E3 are enclosed in square brackets. This applies equally to references to full-entry references and to references to particular senses. In the following example two *cf.* cross-references have been added in E3.

ЖЕНИНСТВО, *n.* Heiratsgut, *dos.* [*cf.* прѣија 1, мираз].

Because both cross-references have been added in E3, the <xr> container receives attribute @source with the corresponding value.

```
<entryFree xml:id="VSK.SR.женинство">
  <form type="lemma">
    <orth norm="женинство">ЖЕНИНСТВО</orth>
  </form>,
  <gramGrp>
    <pos>n.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.женинство.1">
    <cit type="translation">
      <quote xml:lang="de">Heiratsgut</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">dos</quote>
    </cit>.
    [<xr source="#thirdEd" type="cf">
      <lbl>cf.</lbl>
      <ref target="#VSK.SR.прѣија.1">
        <rs type="lemma">прѣија</rs>
        <rs type="senseNum">1</rs>
      </ref>,
      <ref target="#VSK.SR.мираз">
        <rs type="lemma">мираз</rs>
      </ref>
    </xr>].
  </sense>
</entryFree>
```

It may appear at first that the use of <rs> within <ref> is an overstretch. After all, our Encoding Principle 1 (see p. 149) recommends simpler over more complex

solutions; and the content model of <ref> does not prevent us from using text nodes. So why <rs>?

Consider the following example: the entry for *док* has a number of cross-references, including one that points to two numbered senses, instead of the usual one.

до̀к, *bis* (in der Zeit), *usque ad* [cf. 1 *дока*, *доке*, докле 1 и 2, *докале*, *доклен*, *што 7*]: *док* дође; *док* присније; *Док* се једном не смркне, другом не мож' сванути; *Док* се гора преођене листом —

This is a good example in which lexicographic and typographic representations of lexical data are at odds with each other. Fully expanded, the two cross-references would look like this: *докле 1* и *докле 2*, but their abbreviated version (*докле 1 и 2*) makes sense in the context of the limited real estate that the printed page has to offer.

By dividing cross-references into a three-tiered hierarchy consisting of a typed *xr* container on the first level, labels and possibly typed *ref* elements on the second, and segmented and typed *rs* elements on the third level, we can easily encode condensed cross-references as above, remain true to the typographic representation of the historic source, while creating machine-actionable links at the same time.

```
<xr source="#thirdEd" type="cf">
  <!-- etc. -->
  <ref target="#VSK.SR.докле.1">
    <rs type="lemma">докле</rs>
    <rs type="senseNum">1</rs>
  </ref> и
  <ref target="#VSK.SR.докле.2">
```

```

    <rs type="senseNum">2</rs>
  </ref>
  <!-- etc. -->
</xr>

```

5.4.1.5 Cf. with a specific sense numbered in E3

In some cases, however, only the sense number has been added in E3 to an already existing cross-reference from E2.

ждријело, н. (у Херц.)| ber Engraß, fauces: ко
ждрило, н. (зап.) | је витез, потеци на
ждријело (кад се зове у помоћ или у по-
 тјеру). cf. ждрло, богаз [1], кланац [1],
 жвало [2], гротло [2, тјеснац].

In those case @source="#thirdEd" is added to the reference string inside <ref> to show its provenance.

```

<entryFree xml:id="VSK.SR.ждријело">
  <form type="lemma">
    <orth norm="ждријело">ждријело</orth>
    <gramGrp>
      <pos>n.</pos>
    </gramGrp>
    <usg type="diatopic">(у <rs key="912" type="geo">Херц.</rs>)</usg>
  </form>,
  <form type="lemma" xml:id="VSK.SR.ждрило">
    <orth norm="ждрило">ждрило</orth>
    <gramGrp>
      <pos>n.</pos>
    </gramGrp>
    <usg type="diatopic">(<rs type="geo">зап.</rs>)</usg>
  </form>
  <sense xml:id="VSK.SR.ждријело.1">
    <!-- etc. -->
    <xr type="cf">
      <lbl>cf.</lbl>
      <ref target="#VSK.SR.ждрло">
        <rs type="lemma">ждрло</rs>
      </ref>,
      <ref target="#VSK.SR.богаз.1">
        <rs type="lemma">богаз</rs>
      </ref>
    </xr>
  </sense>
</entryFree>

```

```

    <rs type="senseNum" source="#thirdEd">[1]</rs>
  </ref>,
  <ref target="#VSK.SR.кланац.1">
    <rs type="lemma">кланац</rs>
    <rs type="senseNum" source="#thirdEd">[1]</rs>
  </ref>,
  <ref target="#VSK.SR.жвало.2">
    <rs type="lemma">жвало</rs>
    <rs type="senseNum" source="#thirdEd">[2]</rs>
  </ref>,
  <ref target="#VSK.SR.гротло.2">
    <rs type="lemma">гротло</rs>
    <rs type="senseNum" source="#thirdEd">[2</rs>
  </ref>,
  <ref target="#VSK.SR.тјеснац" source="#thirdEd">
    <rs type="lemma">тјеснац</rs>
  </ref>.
</xr>
</sense>
</entryFree>

```

The above example also shows how we've solved the problem of overlapping hierarchies introduced by the typographic conventions used in the print edition: *гротло* [2, *тјеснац*] indicates that that the former reference was present in E2 but without the sense number, which has been added in E3; and that the latter reference was added in E3. The square brackets enclose, however, only the sense number of the first reference and the whole second reference. This is why we have decided to provide lexicographically relevant information at the level of XML attributes, and to make those independent of the typographical representation. Because all our ref elements within <xr> consist of separately encoded reference strings (<rs>), we can make sure to put @source="#thirdEd" at each level of granularity that may be required, i.e. on the <xr> container, the <ref> group or, indeed, individual <rs> elements when necessary.

5.4.1.6 Cf. with homographs

Cross-references pointing to particular homographs have their homograph number preceding the target lemma.

жу́лье́не, n. 1) das Drücken des Schuhs, pressio calcei. [cf. 1 жу́ла́не]. — 2) (у И. г.) vide [чу́па́не] гу́ле́не.

In the above example, *жу́ле́не* refers to the first homograph of *жу́ла́не*. Here too, we encode a computationally actionable link in the target attribute by using the specific homograph id; and we encode the typographic realization of the homograph number as a `<rs type="homNum">` inside the `<ref>` group.

```
<entryFree xml:id="VSK.SR.жу́ле́не">
  <form type="lemma">
    <orth norm="жу́ле́не">жу́ле́не</orth>
  </form>,
  <gramGrp>
    <pos>n.</pos>
  </gramGrp>
  <sense n="1" xml:id="VSK.SR.жу́ле́не.1"><lbl type="senseNum">1)</
    lbl>
    <cit type="translation">
      <quote xml:lang="de">das Drücken des Schuhs</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">pressio calcei</quote>
    </cit>.
    [<xr source="#thirdEd" type="cf">
      <lbl>cf.</lbl>
      <ref target="#VSK.SR.жу́ла́не1">
        <rs type="homNum">1</rs>
        <rs type="lemma">жу́ла́не</rs>
      </ref>
    </xr>].
  </sense>
  <sense n="2" xml:id="VSK.SR.жу́ле́не.2">
    <!-- etc. -->
  </sense>
</entryFree>
```

5.4.1.7 Cf. with a specific sense in a particular homograph

жу́жо, м. hyp. v. жу́жан. [cf. 1 жу́жа 3].

```
<entryFree xml:id="VSK.SR.жу́жо">
  <form type="lemma">
    <orth norm="жу́жо">жу́жо</orth>
  </form>,
  <gramGrp>
    <pos>m.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.жу́жо.1">
    <xr type="hyp">
      <lbl>hyp. <seg xml:lang="de">v.</seg></lbl>
      <ref target="#VSK.SR.жу́жан">
        <rs type="lemma">жу́жан</rs>
      </ref>
    </xr>.
    [<xr source="#thirdEd" type="cf">
      <lbl>cf.</lbl>
      <ref target="#VSK.SR.жу́жа1.3">
        <rs type="homNum">1</rs>
        <rs type="lemma">жу́жа</rs>
        <rs type="senseNum">3</rs>
      </ref>
    </xr>].
  </sense>
</entryFree>
```

5.4.1.8 Cf. changed to vide in E3

When editors changed a cross-reference from cf. to vide in E3, we use a special

@type="cf_to_fide" on <xr>:

же́лезо, н. (око Тимока) đaz Ćižen, ferrum. cf. [vide] гвожђе.

```
<entryFree xml:id="VSK.SR.же́лезо">
  <form type="lemma">
    <orth norm="же́лезо">же́лезо</orth>
  </form>,
  <gramGrp>
    <pos>n.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.же́лезо.1">
    <usg type="diatopic">(око <rs key="1813" type="geo">Тимока</
```

```

    rs>)/usg>
<cit type="translation">
  <quote xml:lang="de">das Eisen</quote>
</cit>,
<cit type="translation">
  <quote xml:lang="la">ferrum</quote>
</cit>.
<xr type="cf_to_vidē">
  <lbl>cf.</lbl>
  <lbl source="#thirdEd">[vidē]</lbl>
  <ref target="#VSK.SR.гвожђе">
    <rs type="lemma">гвожђе</rs>
  </ref>
</xr>.
</sense>
</entryFree>

```

5.4.1.9 Cf. with a foreign word: hidden etymologies

In a small number of cases, cf. is used to refer to foreign words. One may be tempted to encode these as cross-references, but their role in the entry is not to link — even to an imaginary dictionary in a foreign language, but instead to provide etymological information in the context of the current entry. In accordance with our Encoding Principle 3 (p. 149), we have chosen the semantically most concrete TEI element — <etym> — to encode this information.

бања, f. 1) *baš Vab, balneum* (cf. ital. bagno). [cf. топлица 1, илица]. — [*Бања*] 2) варош у Србији (близу Ниша) [сад *Соко-Бања*, у срезу бањском округа нишкога], *Name einer Stadt*,

```

<entryFree xml:id="VSK.SR.бања">
  <form type="lemma">
    <orth norm="бања">бања</orth>
  </form>,
  <gramGrp>
    <pos>f.</pos>
  </gramGrp>
  <sense n="1" xml:id="VSK.SR.бања.1">
    <lbl type="senseNum">1</lbl>
    <cit type="translation">

```

```

    <quote xml:lang="de">das Bad</quote>
  </cit>,
  <cit type="translation">
    <quote xml:lang="la">balneum</quote>
  </cit>
  <etym>
    (<lbl>cf.</lbl>
    <lang value="it">ital.</lang>
    <mentioned xml:lang="it">bagno</mentioned>)
  </etym>.
  <!-- etc. -->
</sense>
<!-- etc. -->
</entryFree>

```

5.4.1.10 Vide

The encoding of vide-synonyms is based on the same principles as the encoding of cf-synonyms. Examples will be listed here for the sake of completeness.

жега, f. vide припека.

```

<entryFree xml:id="VSK.SR.жега">
  <form type="lemma">
    <orth norm="жега">жега</orth>
  </form>,
  <gramGrp>
    <pos>f.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.жега.1">
    <xr type="vide">
      <lbl>vide</lbl>
      <ref target="#VSK.SR.припека">
        <rs type="lemma">припека</rs>
      </ref>
    </xr>.
  </sense>
</entryFree>

```

5.4.1.11 Vide with a specific sense

живо, adv. [ѣхнѣл, velociter. [vide брзо 1].

```

<entryFree xml:id="VSK.SR.живо">

```

```

<form type="lemma">
  <orth norm="живо">жѣво</orth>
</form>,
<gramGrp>
  <pos>adv.</pos>
</gramGrp>
<sense xml:id="VSK.SR.живо.1">
  <cit type="translation">
    <quote xml:lang="de">schnell</quote>
  </cit>,
  <cit type="translation">
    <quote xml:lang="la">velociter</quote>
  </cit>.
  [<xr source="#thirdEd" type="vide">
    <lbl>vide</lbl>
    <ref target="#VSK.SR.брзо.1">
      <rs type="lemma">брзо</rs>
      <rs type="senseNum">1</rs>
    </ref>
    </xr>].
</sense>
</entryFree>

```

5.4.1.12 Vide with homographs

**жнијевати, жнијевâм, v. impf. у пјесми мјесто
жњети [vide 1 жети]: Жнијева је Копчића
робиња | Десном руком, сребрвијем српом —**

```

<entryFree xml:id="VSK.SR.жнијевати">
  <form type="lemma">
    <orth norm="жнијевати">жнијевати</orth>
  </form>
  <!-- etc. -->
  <sense xml:id="VSK.SR.жнијевати.1">
    <usg norm="poetry" type="diatextual">у пјесми мјесто
    <ref target="#VSK.SR.жњети" rend="normal">
      <rs type="lemma">жњети</rs>
    </ref>
  </usg>
  [<xr source="#thirdEd" type="vide">
    <lbl>vide</lbl>
    <ref target="#VSK.SR.жети1">
      <rs type="homNum">1</rs>
      <rs type="lemma">жети</rs>
    </ref>
    </xr>]

```



```

    <!-- etc -->
  </sense>
</entryFree>

```

5.4.1.13 Vide with a specific sense in a particular homograph

Sometimes, both the sense number and the homograph number have been added in E3.

Фѹрати, фѹрѡм, в. ѡмрѡ. vide [2] бацати [1].

```

<entryFree xml:id="VSK.SR.фурати">
  <form type="lemma">
    <orth norm="фурати">фѹрати</orth>
  </form>
  <!-- etc -->
  <sense xml:id="VSK.SR.фурати.1">
    <xr type="vide">
      <lbl>vide</lbl>
      <ref target="#VSK.SR.бацати2.1">
        <rs type="homNum" source="#thirdEd">[2]</rs>
        <rs type="lemma">бацати</rs>
        <rs type="senseNum" source="#thirdEd">[1]</rs>
      </ref>
    </xr>.
  </sense>
</entryFree>

```

5.4.1.14 Vide changed to cf. in E3

If vide from E2 has been changed to cf. in E3, we use @type="vide_to_cf" on <xr>.

2. дѣкаѣе, у загонѣи, vide [cf.] чаре.

```

<entryFree xml:id="VSK.SR.декаѣе2" type="hom" n="2"
  corresp="#VSK.SR.декаѣе1">
  <lbl type="homNum">2.</lbl>
  <form type="lemma">
    <orth norm="декаѣе">дѣкаѣе</orth>
  </form>,
  <sense xml:id="VSK.SR.декаѣе2.1">
    <usg type="diatextual" norm="riddle">у загонѣи</usg>,
    <xr type="vide_to_cf">
      <lbl>vide</lbl>
      <lbl source="#thirdEd">[cf.]</lbl>
    </xr>
  </sense>
</entryFree>

```

```

    <ref target="#VSK.SR.чапе">
      <rs type="lemma">чапе</rs>
    </ref>
  </xr>.
</sense>
</entryFree>

```

5.4.1.15 *Vide changed to dim. in E3*

**же́вка́рити, же́вка́ри̑м, в. | impf. vide [dim. в.]
же́вкати.**

```

<entryFree xml:id="VSK.SR.жевкарити">
  <form type="lemma">
    <orth norm="жевкарити">же́вка́рити</orth>
  </form>,
  <form type="inflected">
    <orth>же́вка́ри̑м</orth>
  </form>,
  <gramGrp>
    <pos>в. impf.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.жевкарити.1">
    <xr type="vide_to_dim">
      <lbl>vide</lbl>
      <lbl source="#thirdEd">dim. <seg xml:lang="de">в.</seg></lbl>
      <ref target="#VSK.SR.жевкати">
        <rs type="lemma">же́вкати</rs>
      </ref>
    </xr>.
  </sense>
</entryFree>

```

5.4.1.16 *Variants*

When *vide* refers to a “correct” variant of the current word, as opposed to a synonym, the target reference appears in bold in the print edition of *LSGL*. In those cases we add @type="variant" to the <ref>.

фа́јда[*], f. vide ва́јда.

```

<entryFree xml:id="VSK.SR.фајда">

```

```

<form type="lemma">
  <orth norm="фајда">фа̀јда</orth>
</form>
<etym source="#thirdEd"><lang value="tr">[*]</lang></etym>
<gramGrp>
  <pos>f.</pos>
</gramGrp>
<sense xml:id="VSK.SR.фајда.1">
  <xr type="vide">
    <lbl>vide</lbl>
    <ref target="#VSK.SR.вајда" type="variant">
      <rs type="lemma">вајда</rs>
    </ref>
  </xr>.
</sense>
</entryFree>

```

Reversely, when an entry is pointing to a less preferred variant, the variant appears in regular font next to the lemma. We treat these as implicit *cf* cross-references (<xr type="cf">) based on the distinction between *vide* and *cf*, as described above in Section 5.4.1 on p. 187, while the <ref> is given @type="variant".

ва̀јда* [фа̀јда], f. Nutzen, utilitas. cf. [vide] ко-
рист.

```

<entryFree xml:id="VSK.SR.вајда">
  <form type="lemma">
    <orth norm="вајда">ва̀јда</orth>
  </form>
  <etym>
    <lang value="tr">*</lang>
  </etym>
  [<xr type="cf" source="#thirdEd">
    <ref target="#VSK.SR.фајда" type="variant">
      <rs type="lemma">фајда</rs>
    </ref>
  </xr>]
  <!-- etc. -->
</entryFree>

```

5.4.2 *Implicit cross-references*

By implicit cross-references, I refer to cross-references in the print edition that lack an explicit label for the type of referring (*cf.*, *vide*, *vidu s. v.* and so forth). Implicit cross-references appear in the regular flow of text (inside definitions or narrative sections in the entry (<dictScrap>, see Section 5.3.6.4, pp. 180-181) are recognized by their typographic features (cursive or extended font) and their context. For implicit references, we do not use the <xr> container element since its role could be reduced to grouping together references with their labels. In the absence of labels, <xr> can be safely omitted.

5.4.2.1 *Cursive references*

Cursive has been used in E1 and E2 to refer to related words and not only synonyms. These cursive references have been preserved in E3 and beyond. In the entry *женидба* [wedding], which contains long narrative sections about the customs associated with the wedding ceremony, we encounter a sentence like this:

У сватовима мора бити кум, ђевер, стари сват, прикумак, војвода, чауш, и гадљар; а остали сви се зову пустосватице, или (као у шали) набигузике. Кума треба да зовне

Loosely translated, this sentence reads:

The wedding party must include *kum* [best man], *đever* [brother of the groom], *stari svat* [witness], *prikumak* [best man's helper], *vojvoda* [groom's uncle on the maternal side], *čauš* [master of ceremonies] and *gadljär* [bagpiper]; and the others are called *pustosvatice* [wedding guests without a specific role], or (jokingly) *nabiguzice* [freeloaders, lit. those who stuff their asses].

and is encoded in XML like this:

```
<dictScrap>
  <!-- etc. -->
  У сватовима мора бити
  <ref target="#VSK.SR.кум" rend="italic">
    <rs type="lemma">кум</rs>
  </ref>,
  <ref target="#VSK.SR.ђевер" rend="italic">
    <rs type="lemma">ђевер</rs>
  </ref>,
  <ref target="#VSK.SR.стари_сват" rend="italic">
    <rs type="lemma">стари сват</rs>
  </ref>,
  <ref target="#VSK.SR.прикумак" rend="italic">
    <rs type="lemma">прикумак</rs>
  </ref>,
  <ref target="#VSK.SR.војвода.5" rend="italic">
    <rs type="lemma">војвода</rs>
  </ref>,
  <ref target="#VSK.SR.чауш" rend="italic">
    <rs type="lemma">чауш</rs>
  </ref>,и
  <ref target="#VSK.SR.гадљар" rend="italic">
    <rs type="lemma">гадљар</rs>
  </ref>; а остали сви се зову
  <ref target="#VSK.SR.пустосват" rend="italic">
    <rs type="lemma">пустосватице</rs>
  </ref>, или (као у шали)
  <ref target="#VSK.SR.набигузица" rend="italic">
    <rs type="lemma">набигузице</rs>
  </ref>.
  <!-- etc. -->
</dictScrap>
```

5.4.2.2 *Extended references*

There are two types of implicit references that the editors of E3 highlighted with an extended font: (a) words that were not marked up as implicit references in E2 but were added in E3; and (b) words inside monolingual definitions or narrative sections that did not have an entry of their own in E2. In both cases these references get an additional attribute `@rend="extended"`.

New implicit references

Implicit references introduced in E3 were typographically set apart using an extended font so that they could be distinguished from the already present implicit references. For instance, if we compare the entry for баба in E2:

БАБА, f. (gen. pl. баба̂) 1) очина или материна мајка, die Großmutter, avia. 2) стара жена, altes Weib, vetula. 3) у Дубровнику се свака дојкиња и дадиња зове баба, макар била и дјевојка од 12 година, die Amme, Kindsfrau,

with its corresponding entry in E3:

1. баба, f. (gen. pl. баба̂) 1) очина или материна мајка, die Großmutter, avia. [cf. стара 2]. — 2) стара жена, altes Weib, vetula. [cf. стара 1, старица]. — 3) у Дубровнику се свака дојкиња и дадиња зове баба, макар била и дјевојка од 12 година, die Amme, Kindsfrau,

we'll notice that in the definition under sense no. 3, in both E2 and E3 баба was a self-reference (i.e. an instance of the lemma in the text of the dictionary); whereas дојкиња and дадиња were regular words in E2, but were turned into implicit reference in E3.

```
<def>y <rs type="geo" key="270">Дубровнику</rs> се свака
<ref target="#VSK.SR.дојкиња" rend="extended">
  <rs type="lemma">дојкиња</rs>
</ref> и
<ref target="#VSK.SR.дадиња" rend="extended">
  <rs type="lemma">дадиња</rs>
</ref> зове
<hi rend="italic">баба</hi>, макар била и дјевојка од 12
година
</def>
```

Implicit references for new entries

The editors of E3 used an extended font also to mark up words inside entries that did not have their own entry in E2, but which were subsequently added in E3. For instance, the definition for *заклад* reads „лијепе хаљине, оружје и остале драгоцене ствари” [beautiful attire, weapons and other precious things]. The adjective *драгоцјене* [precious, nom. fem. pl.] is typographically set apart from the rest of the entry:

за́клад, м. (по јужн. кр.) 1) лијепе хаљине, оружје, и остале драгоцене ствари, даџ Kleinod, ornamentum. [cf. заклада]. — 2) даџ Pfand, pignus [vide залог]. cf. залога.

This is because in E2 there was no entry for the adjective *драгоцјен*, but it has been added in E3 (and marked with square brackets).

[драгоцјен, а, о, vide скупоцјен; види s. v. заклад.]

This type of cross-references receives an additional attribute `@rend="extended"`:

```
<entryFree xml:id="VSK.SR.заклад">
  <form type="lemma">
    <orth norm="заклад">за́клад</orth>
  </form>,
  <gramGrp>
    <pos>м.</pos>
  </gramGrp>
  <usg type="diatopic">(по <rs type="geo">јужн. кр.</rs>)</usg>
  <sense n="1" xml:id="VSK.SR.заклад.1">
    <lbl type="senseNum">1</lbl>
    <def>лијепе хаљине, оружје, и остале <ref
      target="#VSK.SR.драгоцјен" rend="extended"><rs
        type="lemma">драгоцјене</rs></ref> ствари</def>
    <!-- etc. -->
  </sense>
  <sense n="2" xml:id="VSK.SR.заклад.2">
```

```

    <!-- etc. -->
  </sense>
</entryFree>

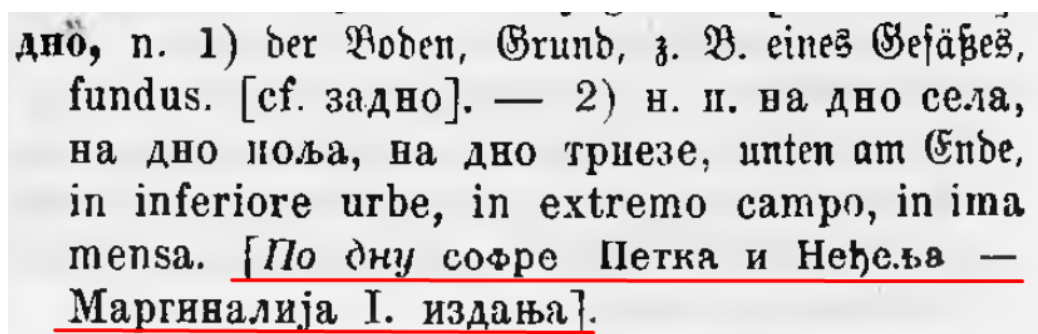
```

5.5 Editorial interventions

E3 (and E4 which is identical to E3) is a lexicographic palimpsest, containing the contents of E1 and E2 (with some exceptions, as we'll see later), Vuk's handwritten marginalia from his personal copies of E1 and E2, as well as a series of editorial interventions which were introduced in E3.

5.5.1 Marginalia from E1

Vuk's annotations from the marginalia of his copy of E1 are added in E3 in square brackets with a note „Маргиналија I. издања” [First edition marginalia]. For instance:



дно, n. 1) der Boden, Grund, z. B. eines Gefäßes, fundus. [cf. задно]. — 2) н. п. на дно села, на дно поља, на дно трпезе, unten am Ende, in inferiore urbe, in extremo campo, in ima mensa. [По дну софре Петка и Неђеља — Маргиналија I. издања].

In cases such as these, it was important to separate the actual content of the marginalia from the editor's note. The original content is then treated as part of the dictionary hierarchy, with an appropriate source attribute to identify its provenance, whereas the editor's note is marked up as the editor's addition in E3:

```

<entryFree xml:id="VSK.SR.дно">
  <form type="lemma">
    <orth norm="дно">дно</orth>
  </form>,
  <gramGrp>

```



```

    <pos>n.</pos>
  </gramGrp>
  <!-- etc. -->
  <sense n="2" xml:id="VSK.SR.дно.2">
    <lbl type="senseNum">2</lbl>
    <usg type="hint">н. п. на дно села</usg>,
    <usg type="hint">на дно поља</usg>,
    <usg type="hint">на дно трпезе</usg>,
    <cit type="translation">
      <quote xml:lang="de">unten am Ende</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">in inferiore urbe</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">in extremo campo</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">in ima mensa</quote>
    </cit>.
    [<cit type="example" source="#firstEd-marginalia">
      <quote xml:lang="sr"><hi rend="italic">По дну</hi> софре Петка
        и Неђеља</quote>
    </cit> –
    <note source="#thirdEd">Маргиналија I. издања</note>].
  </sense>
</entryFree>

```

5.5.2 Accidental omissions from E2

Entries that were accidentally omitted from E2 have been added in E3 in square brackets and an editorial note „Из I. издања” [From the first edition]:

**[накѣњити се, им се, v. г. pf. њѣ ѡтрапатиен,
fatigor. — Из I. издања.]**

Here, too, as in the case of E1 marginalia, we are using @source="#firstEd" to highlight the origin of the element it is attached too, while we source the editor's note as @source="#thirdEd" because this note was added in E3. To distinguish entries that were accidentally omitted from those that were intentionally removed

(see Section 5.5.4 Vulgarisms from E1 [pp. 215-216]), we're adding

@type="accidentallyOmittedFromSecondEd" to the former.

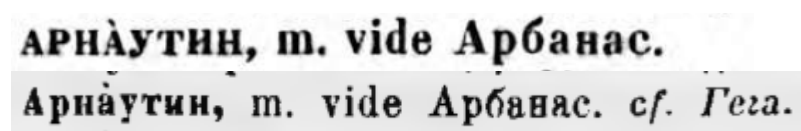
```
<entryFree xml:id="VSK.SR.накињити_се" source="#firstEd"
  type="accidentallyOmittedFromSecondEd">
  [<form type="lemma">
    <orth norm="накињити се">накињити се</orth>
  </form>,
  <form type="ending">
    <orth>ѣм се</orth>
  </form>,
  <gramGrp>
    <pos>v. r. pf.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.накињити_се.1">
    <cit type="translation">
      <quote xml:lang="de">sich strapazieren</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">fatigor</quote>
    </cit>.
  </sense>
  <note source="#thirdEd">— Из I. издања.</note>]
</entryFree>
```

5.5.3 Marginalia from E2

E3 introduced two types of interventions from Vuk's annotations of E2: additions to existing entries, and completely new entries. In both cases, the editors of E3 used a special type of cursive to distinguish those interventions from the regular text.

5.5.3.1 Additions to existing entries

Consider the following comparison of the entry *арнауџин* in E2 and E3:



АРНАУТИН, m. vide Арбанас.
 Арнауџин, m. vide Арбанас. cf. Геза.

The cursive styling of cross-reference (*cf. Геза*) indicates that this cross-reference was suggested by Vuk on the margins of his copy of E2. In the digital edition, this is

recorded with a corresponding @source="#secondEd-marginalia" and a @rend="italic".

```
<entryFree xml:id="VSK.SR.Арнаутин">
  <form type="lemma">
    <orth norm="Арнаутин">Арна̀утин</orth>
  </form>,
  <gramGrp>
    <pos>m.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.Арнаутин.1">
    <xr type="vide">
      <lbl>vide</lbl>
      <ref target="#VSK.SR.Арбанас">
        <rs type="lemma">Арбанас</rs>
      </ref>
    </xr>.
    <xr rend="italic" source="#secondEd-marginalia" type="cf">
      <lbl>cf.</lbl>
      <ref target="#VSK.SR.Гера">
        <rs type="lemma">Гера</rs>
      </ref>
    </xr>.
  </sense>
</entryFree>
```

The italicization of certain elements leads to a typographic reversal for self-references and implicit cross-references (see Section 5.4.2 [pp. 204-208]), which would otherwise be themselves italicized.

Let us, for instance, consider the entry *камен* [stone] as it appears in E2 and its counterpart, 1. *камен*, in E3:

**КАМѢН, камена, т. (loc. камѣну) der Stein, lapis,
cf. кам, камн: Од камена ништа до ка-
мена.**

1. **КАМѐН**, **КАМЕНА**, **т.** (лос. **камѐну**) 1) *der Stein, lapis, cf. кам, камн [крш 2]: Од камена ништа до камена.* — 2) *(у Ц. г.) отишао на свој камен, т. ј. на своју постојбину; тако се каже и: Не ће да остави својега камена, и т. д. [— 3) в. п. воденица од три камена. cf. ВИТАО 3].*

We can observe the following:

1. E2 entry contained a single unnumbered sense which became a numbered sense in E3;
2. sense numbers for senses 1 and 2 in E3 appear in italics, which means that they were added by Vuk in his notes on the E2;
3. the contents of sense no. 2 in E3 are italicized with the exception of „на свој камен”, „постојбину” and „својега камена”; this is an example in which the dictionary’s typographical conventions have been reversed;
4. sense no. 3 was added by the editors in E3 and is hence marked off by square brackets.

In a literal translation, sense no. 2 would read:

2) *(in Montenegro) he went to his rock, i.e. to his home(land); one also says: He doesn’t want to leave his rock, etc.*

The first and the third instanced of reversed cursive is self-referential: the syntagm *свој камен* [one’ rock, one’s stone] is treated as a reference to the headword of this entry. *Rock* in this sense does not denote a solid mineral material (which is covered by sense no. 1). Instead, *one’s rock* is a metaphoric expression describing the place

one is form. Hence the second instance of reversed cursive in the above example is an implicit cross-reference to *ѡсѡѡјдина* [place of origin, homeland].

```
<entryFree xml:id="VSK.SR.камен1" type="hom" n="1"
  corresp="#VSK.SR.камен2">
  <lbl type="homNum" source="#thirdEd">1.</lbl>
  <form type="lemma">
    <orth norm="камен">кămêн</orth>
  </form>,
  <form type="inflected">
    <orth>кăмeнa</orth>
  </form>,
  <gramGrp>
    <pos>m.</pos>
  </gramGrp>
  (<form type="inflected">
    <lbl>loc.</lbl>
    <orth>кaмêнy</orth>
  </form>)
  <sense n="1" xml:id="VSK.SR.камен1.1">
    <lbl type="senseNum" rend="italic" source="#secondEd-
      marginalia">1)</lbl>
    <cit type="translation">
      <quote xml:lang="de">der Stein</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">lapis</quote>
    </cit>,
    <xr type="cf">
      <lbl>cf.</lbl>
      <ref target="#VSK.SR.кам">
        <rs type="lemma">кам</rs>
      </ref>,
      <ref target="#VSK.SR.ками">
        <rs type="lemma">ками</rs>
      </ref>
      [<ref target="#VSK.SR.крш.2">
        <rs type="lemma">крш</rs>
        <rs type="senseNum">2</rs>
      </ref>,
      <ref target="#VSK.SR.стијена.3">
        <rs type="lemma">стијена</rs>
        <rs type="senseNum">3</rs>
      </ref>]
    </xr>:
    <cit type="example">
      <quote xml:lang="sr">Од <hi rend="italic">камена</hi> ништа до
        <hi rend="italic">камена</hi>.
      </quote>
```

```

</cit>
</sense>
<sense n="2" source="#secondEd-marginalia"
  xml:id="VSK.SR.камен1.2" rend="italic">
  <lbl type="senseNum">2</lbl>
  <usg type="diatopic">(у <rs key="930" type="geo">Ц. г.</rs></
  usg>
  <cit type="exmple">
    <quote xml:lang="sr">отишао <hi rend="regular">на свој камен</
    hi></quote> <usg type="hint">т. ј. на своју <ref
    target="#VSK.SR.постојбина" rend="regular">постојбину</ref></
    usg>
  </cit>; тако се каже и:
  <cit type="example">
    <quote xml:lang="sr">Не ће да остави <hi
    rend="regular">својега камена</hi>
    </quote>
  </cit>, и т. д.
</sense>
<sense n="3" source="#thirdEd" xml:id="VSK.SR.камен1.3">
  <lbl type="senseNum">3</lbl>
  <cit type="example">
    <usg type="hint">н. п.</usg>
    <quote xml:lang="sr">воденица од три <hi
    rend="italic">камена</hi>
    </quote>
  </cit>.
  <xr type="cf">
    <lbl>cf.</lbl>
    <ref target="#VSK.SR.витао.3">
      <rs type="lemma">витао</rs>
      <rs type="senseNum">3</rs>
    </ref>
  </xr>.
</sense>
</entryFree>

```

5.5.3.2 New entries

The same principle of italicization that we have seen in the previous section applies to entire entries that Vuk jotted down on the margins of the published E2. For instance:

*блѣнда, ф. (у Бан. и у Бачк.) као шклопац,
н. и. кад ујезде комарац.*

```

<entryFree xml:id="VSK.SR.бланда" source="#secondEd-marginalia"
  rend="italic">
  <form type="lemma">
    <orth norm="бланда">блâнда</orth>
  </form>,
  <gramGrp>
    <pos>f.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.бланда.1">
    <usg type="diatopic">(у <rs key="48" type="geo">Бан.</rs> и у
      <rs key="57" type="geo">Бачк.</rs>)</usg>
    <def>као шклопац, н. п. кад уједе комарац</def>.
  </sense>
</entryFree>

```

5.5.4 *Vulgarisms from E1*

The digital edition of *LSGL* is the first edition of Vuk's dictionary since 1818 which contains all the vulgarisms (both fully and partially vulgar entries) that Vuk intentionally removed from E2 under pressure from the conservative elites and the Serbian Orthodox Church (Селимовић 1967; Леовац 1987). Because of the conservative editorial policies of the largest dictionary of the Serbian language (САНУ 1959), some of these words were never lexicographically treated after Vuk. Vulgar content has been extracted from E1 manually and its place in both the dictionary macro- and microstructure has been reconstructed. To identify this type of content, we have used `@source="#firstEd"` and `@type="intentionallyOmittedFromSecondEd"` in entries, or `@subtype="intentionallyOmittedFromSecondEd"` in examples.

For instance, *кривокуп* [adj. having a curved penis]:

```

<entryFree xml:id="VSK.SR.кривокуп" source="#firstEd"
  type="intentionallyOmittedFromSecondEd">
  <form type="lemma">
    <orth norm="кривокуп">кривòкуп</orth>
  </form>,
  <form type="ending">
    <orth>па</orth>
  </form>,
  <form type="ending">

```

```

    <orth>по</orth>
  </form>,
  <gramGrp source="#digitalEd">
    <pos>adj.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.кривокуп.1">
    <cit type="translation">
      <quote xml:lang="la">penis incurvi</quote>
    </cit>.
    <cit type="example">
      <quote xml:lang="sr">Има и као и кривокури Татар</quote>
    </cit>.
  </sense>
</entryFree>

```

or the vulgar example in *глогиња* [hawthorn]: *Ласно је туђим курцем глогиње млати* [lit. it is easy to beat the hawthorn bush with somebody else's dick, i.e. it's easy to be brave from a distance]:

```

<entryFree xml:id="VSK.SR.глогиња">
  <form type="lemma">
    <orth norm="глогиња">глòгиња</orth>
  </form>,
  <gramGrp>
    <pos>f.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.глогиња.1">
    <cit type="translation">
      <quote xml:lang="de">die Frucht des Weissdorns</quote>
    </cit>,
    <cit type="translation">
      <quote xml:lang="la">fructus crataegi</quote>
    </cit>.
    <cit type="example" source="#firstEd"
      subtype="intentionallyOmittedFromSecondEd">
      <quote xml:lang="sr">Ласно је туђим к***ем глогиње млатити</
      quote>
    </cit>.
  </sense>
</entryFree>

```


IMPLEMENTING AN API-CENTRIC DIGITAL EDITION

6.1 Text as service

As a community of practice, digital humanists deal with texts as complex, multi-dimensional and multi-layered datasets that can be analyzed, annotated and manipulated in order to arrive at new knowledge. In the previous chapters, I have argued that the functional affordances of a digital edition will depend to a large extent on the level of detail provided by the textual markup of the source. If one marks up German and Latin translations as separate structural units in a Serbian-German-Latin dictionary, it will be comparatively easy to create a reverse index for the given dictionary, or let users search for Serbian lemmas based on their translation equivalents. The more (encoding, metadata, annotation) we put into our source documents, the more specific will be the types of data or information that users can retrieve from our digital editions.

Thus far, the use of semantic markup, most commonly according to the Guidelines of the Text Encoding Initiative, has been *the* major focus of digital humanists designing and publishing electronic editions. The benefits of focusing on

structure and semantics rather than the edition's visual aspect during the encoding process have been widely acknowledged. Text encoding is, after all, a form of text analysis. But an encoded text is merely a data source for a digital edition, not the edition itself. This is an important point to keep in mind, because we often tend to let the important difference between the two recede into the background. Only experts can work directly with texts in XML, read them, analyze them and run various tools on them, but the majority of end users need a graphical user interface for interacting with texts.

One of the most consistently recognized features of electronic textuality has been the separation of form and content: the fact that what we see on the screen is different from the code that produced the text. A digital text, in that sense, is a projection of the “actual” text onto a surface:

Die Einheit von Text und Oberfläche ist damit aufgehoben und damit auch der Objektcharakter des digitalen Textes. (Bittner 2003)

The unity of text and surface is thus abolished and with it also the character of the text as *object*. [emphasis mine]

But if digital text is not an object, what is it? Needless to say, that depends on the context and the framework of analysis. It can be seen as a “performance” (Brown and Duguid 1996), a “process” (Hayles 2003), or, in the context of the ongoing discussion of digital editions and text-delivery mechanisms — a *service*: a kind of text that is decoupled from its presentation surface and is delivered over standard protocols as a stream of structured data. I have argued elsewhere (Tasovac 2017) that texts in digital libraries should be machine-computable and not only machine-deliverable; a kind of service as opposed to fixed objects; and an interactive method instead of a static method, but for this disruptive re-thinking of texts *and* libraries to

be possible, the way we conceptualize, implement and disseminate our digital editions needs to change too.

The goal of this chapter is not to discuss the details regarding the engineering aspects of Service Oriented Architectures (SOA) (see, for instance, Erl *et al.* 2017). Instead, I want to argue for the need to recalibrate the traditional production workflow of TEI-based editions. In a prototypical setup, a digital edition will consist of source documents in XML, a set of XSLT stylesheets and an HTML output. In such setups we transform source data in order to project them onto a reading surface. It is considered good practice in the DH community for online digital editions to make their source documents available as well, but more often than not user-interaction with the text, including, most importantly, search functionalities, is inevitably coupled with the user interface.⁶⁹

What I am proposing instead — and what I have achieved with the digital edition of LSGS and other dictionaries on raskovnik.org — is to introduce one additional, crucial, step in the digital edition workflow: the creation of an interactive text delivery mechanism which is independent of the user interface. This is different from simply providing download links to source files in XML. I am arguing for the need to create Application Programming Interfaces (APIs) to the

⁶⁹ A static digital edition is modeled largely on print-based textual forms (Siemens *et al.* 2009) and the “layout economics invented for the printed page” (Vanhoutte 2009: 109). This type of objectification embodies a quite conventional understanding of textuality. While generations of critical theorists showed us that a text is neither self-evident nor self-identical, our digital editions seem to convey the opposite. The most common interface architectonics is still constructed around the notion of a static, stable text at the core of the edition. Like the venerable facsimiles of the print tradition, the source text displayed on the screen is usually immutable. Relegated to the digital periphery, to off-screen electronic margins, the digital paratext remains deeply steeped in the typotopographic conventions of its print counterpart, reinforce both the stasis and status of the main text; upholding its authority and keeping it uncontaminated by external interventions.

data contained in our digital editions in addition to creating user interfaces. And not only for dictionaries, but digital editions in general.

One of the most important challenges facing the DH community today is how to consolidate and repurpose available data and tools; how to create reusable but flexible workflows; and, ultimately, how to integrate and disseminate knowledge, instead of merely capturing it and encapsulating it. This technical and intellectual shift can be seen as the “infrastructural turn” in digital humanities (Tasovac *et al.* 2015). Infrastructures in general are installations and services that function as “mediating interfaces” or “structures ‘in between’ that allow things, people and signs to travel across space by means of more or less standardized paths and protocols for conversion or translation” (Badenoch and Fickers 2010: 11). Digital research infrastructures are no different: they are a mediating set of technologies for research and resource discovery, collaboration, sharing and dissemination of scholarly output.

In the context above-mentioned challenge of integrating available DH resources, Application Programming Interfaces (APIs) have the potential to be powerful, practical building blocks of DH infrastructures. On the technical level, they let heterogeneous agents dynamically access and reuse the same sets of data and standardized workflows. On the social level, they help overcome the problem of “shy data”, i.e. data you can “meet in public places but you can’t take home with you” (Cooper 2010: unpaginated). More than years ago, Dan Cohen started the conversation about APIs in DH by pointing out that, despite their potential, few humanities projects — in contrast to those in the sciences and commercial realms — were developing APIs for their resources and tools (Cohen 2005). In the decade since, API development in the digital humanities has certainly increased: today,

both large-scale, national and international initiatives, such as HathiTrust, DPLA or Europeana, as well as individual projects, such as International Image Interoperability Framework (IIIF)⁷⁰, Canonical Text Services (CTS)⁷¹, Open Siddur,⁷² Folger Digital Texts,⁷³ correspSearch,⁷⁴ Old Bailey,⁷⁵ European Holocaust Research Infrastructure⁷⁶ etc., have shifted the focus of their attention and resources to developing APIs.

My approach to creating API-centric digital editions offers tangible interfaces to textual data that can be used in tailor-made workflows by humanities researchers and other users, well-suited to distant reading techniques, statistical analysis and computer-assisted semantic annotation. This approach, in which an API is the core of the digital edition, and the interface something that is built on top of the data model and the API, does not only have practical implications for the way we build tools and create resources, but also has theoretical ramifications for the way we distinguish highly from loosely structured data: if text is not an object, but a service; and not a static entity, but an interactive method; a formal, computational distinction between a dictionary and, say, a novel or a poem, is more difficult to maintain.

⁷⁰ <http://iiif.io>

⁷¹ <http://www.homermultitext.org/hmt-docs/cite/cite-overview.html>

⁷² <https://github.com/opensiddur/opensiddur/wiki/Intro-to-hacking-the-API>

⁷³ <http://www.folgerdigitaltexts.org/api>

⁷⁴ <http://correspsearch.net/index.xql?id=api&l=en>

⁷⁵ <https://www.oldbaileyonline.org/static/API.jsp>

⁷⁶ <https://www.manuscriptsonline.org/api>

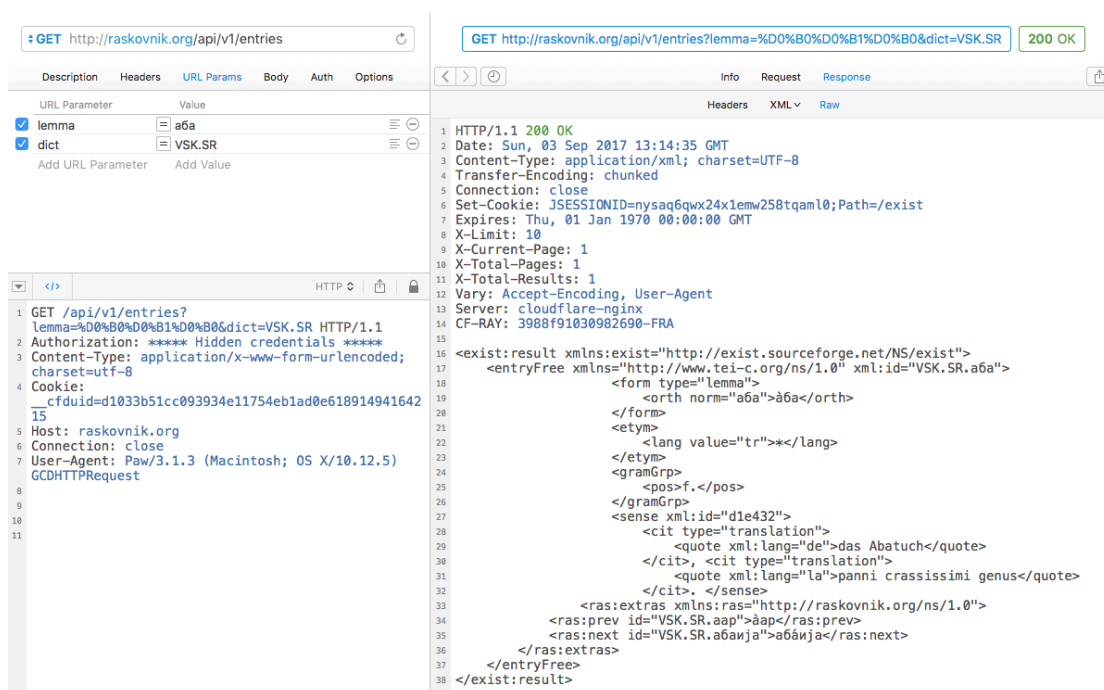
6.2 Raskovnik API: accessing lexical data

Raskovnik API is an application programming interface designed for accessing structured lexical data. I have designed the API in the context of the previously mentioned platform for digital editions of legacy dictionaries that was developed as a cooperation between the Institute of the Serbian Language of the Serbian Academy of Arts and Sciences and the Belgrade Center for Digital Humanities (<http://raskovnik.org>). The interactive API documentation is available online under <http://docs.raskovnikapi.apiary.io>, and a static output of the documentation is also attached as an appendix to this dissertation.

I will not repeat here all the technical details that have been described in the API documentation. Instead I will focus on the thinking behind the API design and how it corresponds to the *functional affordances* of the dictionary as a text-bearing object which I discussed in Chapter 1 and especially in Section 1.3.4 (pp. 78-83). In other words, I will describe how the design of the Raskovnik API can be used as a building block for creating online interfaces that facilitate searching, browsing as well as serendipitous discovery.

6.2.1 Searching

Searching, as we have already seen, is an active directed information-related behavior. The user knows exactly what word he or she is seeking information about. Raskovnik API provides an `/entries` endpoint which returns entries based on the lemma search. For instance a GET request to <http://raskovnik.org/api/v1/entries?lemma=a6a&dict=VSK.SR> will return the corresponding entry from *LSGL*:

Illustration 1: Searching for entry *a6a* via the Raskovnik API

Various parameters can be used to modify the search:

- the parameter **dict** can be used to modify the pool of dictionaries to be considered in the search; **dict=all** (which is the default value) will return entries with a given lemma in all the dictionaries; it is, however, possible to select individual dictionaries by submitting their ids as comma-separated values, for instance: **dict=VSK.SR,MZ.RGJS**.
- the parameter **contains** is used to modify the type of string matching; **contains=false** (the default) will return entries whose lemma is the exact match of the queried string; **contains=true** will return entries whose lemma contains the queried string (for instance, in multiword expressions)
- the parameter **limit** controls the maximum numbers of resources to be returned per page

- the parameter `transform` controls whether the resources will be returned in XML or already transformed to HTML, etc.

In addition to returning the TEI-encoded entry, the API provides, with each entry, links to the immediately preceding and immediately following entry in the given dictionary. So, for the entry *αδα* in *LSGL*, the API will also return:

```
<ras:extras xmlns:ras="http://raskovnik.org/ns/1.0">
  <ras:prev id="VSK.SR.aap">αα</ras:prev>
  <ras:next id="VSK.SR.αβαιja">αβαιja</ras:next>
</ras:extras>
```

The immediate macrostructural context is meant to facilitate the creation of navigation links in the user interface.s

6.2.2 *Browsing*

Raskovnik API recognizes the need for active undirected modes of information seeking. Since our dictionary data is highly structured, the API has been designed in such a way to enable the extraction of dictionary data, both as lemma lists (which can be used for the creation of browsing interfaces) or as full-entries (which can be used to display portions of the dictionary based on the input query). In both cases, the API provides an interface to the dictionary data via XPath expressions (Kay 2008). XPath is a W3C recommended language which allows the addressing of hierarchical nodes in an XML tree (and, as of version 3.1 also JSON) (see Robie *et al.* 2017). The API endpoint which receives XPath expressions as query input is called */slices*. A dictionary “slice” is a list of entries that share a certain lexicographically identifiable and computationally addressable feature. In the context of the Raskovnik API, a slice is a portion of the dictionary macrostructure that has been returned by a specific XPath expression. This technically simple

Thanks to the fact that it is based on XPath expressions, `/slices` is a very powerful and flexible point of entry to exploring dictionary data. It does presuppose, however, that the client is familiar with the structure of the dictionary to be queried. That is the price that has to be paid for having direct access to highly structured and complex data.

6.2.3 *Serendipitous discovery*

Raskovnik API facilitates serendipitous discovery in multiple ways. We have already seen that the `/entries` point returns, for instance, the full text of the queried entry together with the references to the previous and following links. The API, however, also has two dedicated endpoints that can contribute to the discovery process.

The `/context` endpoint returns the siblings of a given entry wrapped in `divs` (with `@type` attribute having possible values `left`, `current` and `right` for the preceding, current and following siblings respectively. The parameters `left` and `right` control the number of entries to be returned in each context; whereas the parameter `return` (with possible values: `left`, `right`, `both`) controls which contexts will be returned. `/range` is similar to `/context` in that it works on the relation by macrostructural proximity, but it does so from a different angle: it returns a span of entries given two individual entry ids. Finally, `/random` is the endpoint which, as the name suggests, selects an entry to return by chance, although the parameter `dict` can filter out or in individual dictionaries to be considered for randomization.

With the Raskovnik API, I wanted to create a data source which could be, on the one hand, plugged in into the frontend of the Raskovnik Dictionary Platform; and, on the other hand, used as an independent web-service. For example: my team at the Belgrade Center for Digital Humanities and I are currently working on a

publication platform for API-centric digital editions of TEI-encoded literary and historical texts called *Izdanak*. Thanks to the conceptual and architectural decisions with *LSGL* and other dictionaries, it will be possible, for example, to use Raskovnik API to query dictionaries directly from <http://izdanak.org> with user-friendly look-up mechanisms: clicking on any word in the text will bring up a pop-up window with dictionary entries for the given word. The reverse case should also be true. When we complete the *Izdanak* API, providing access to the literary and historical texts that we publish, we will be able to retrieve authentic examples from those texts and use them to enrich the existing dictionary entries on <http://raskovnik.org>.

6.3 GUI: affordances in action

The GUI of the digital *LSGL* and other dictionaries on raskovnik.org is based on the insights gained from the analysis of the *functional affordances* of both print and online dictionaries. As I already pointed out in Section 1.3 (pp. 69-83), print dictionaries have traditionally fared better in terms of affording *contextualized reading* because of the way multiple entries are laid out on two opposing pages. Online dictionaries, on the other hand, tend to decontextualize full-text entries by retrieving individual search results and displaying them in isolation. Dictionaries, however, especially those that we have inherited from the past, are not just tools, but also historical documents. User interfaces which provide only database-centric views reduce the user's engagement with historical dictionaries to targeted data-extraction fail to facilitate different types of information-related behavior that are made possible by the dictionary as a complex socio-cultural assemblage.

It is important to emphasize here that an online user interface developed in careful consideration of the functional affordances of the dictionary as such does not necessarily lead to skeuomorphic design.⁷⁷ Quite the contrary. There is a clear line separating the dictionary's functional affordances (which are *general*, typical to the genre as a whole, and related to the various possibilities of seeking and being exposed to information in a dense, alphabetically ordered text-bearing object) from the concrete, visual, graphical or ornamental aspects of *particular* design implementations. As I will show in the rest of this section, the visual design of the digital *LSGL* is minimalist and in no way an attempt to replicate the particular look and feel of the print edition. If one wanted to create a digital surrogate, a facsimile edition would do the job.

6.3.1 Searching

There are two types of lemma searches available on raskovnik.org: a *primary* search on the home page which facilitates searching across all or selected dictionaries hosted on the platform; and a *secondary* search which performs queries on a single dictionary from within the entry display interface. Both primary and secondary search interfaces autocomplete the user's input with a dropdown menu of lemmas that match the given input.

6.3.2 Primary search

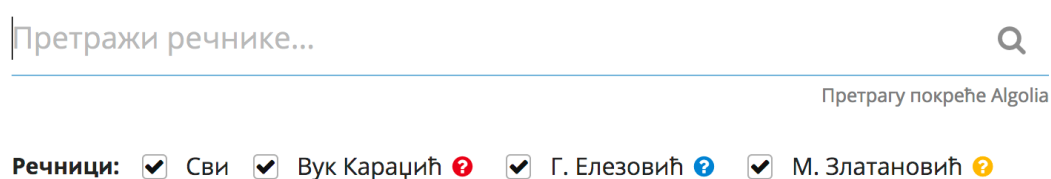
To search for lemmas in *LSGL*, the user enters a query in a search input field on the home page of raskovnik.org. By default the search is performed across all the

⁷⁷ In the archeological context, where the term and the concept were first discussed, a skeumorph is defined as "an element of design or structure that serves little or no purpose in the artifact fashioned from the new material but was essential to the object made from the original material" (Basalla 1988: 107). Nowadays, skeuomorphic design

dictionaries available on the platform, but it is also possible to select one or more dictionaries for the search by checking the checkboxes below the input field (see Illustration 3 below).

Лепа реч гвоздена врата отвара

Откријте и истражите богатство српске лексике



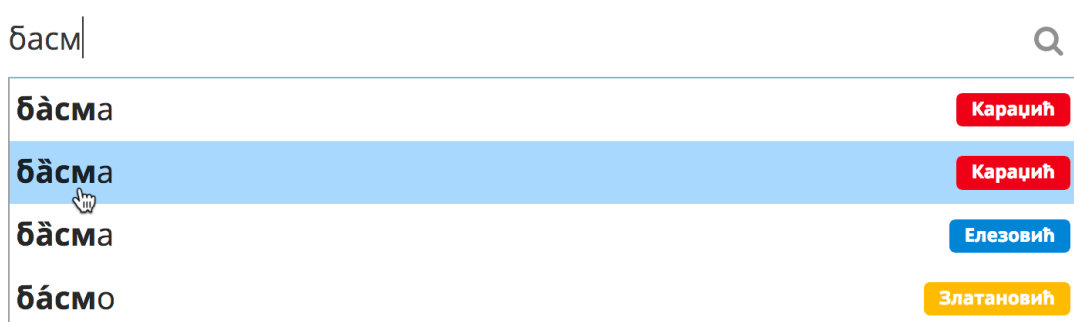
Претражи речнике...

Претрагу покреће Algolia

Речници: ☒ Сви ☒ Вук Караџић ☒ Г. Елезовић ☒ М. Златановић

Illustration 3: Search input field

After entering at least two letters, an autosuggest dropdown menu appears presenting an overview of the available lemmas matching the current input. With each new letter entered, the displayed entries will change to match the changed input. Entries are tagged with the dictionary author's last name for easy identification of sources.



басм

бàсма	Караџић
бàсма	Караџић
бãсма	Елезовић
бáсмо	Златановић

Illustration 4: Search autosuggest dropdown

6.3.3 Secondary search

As far as the Raskovnik GUI is concerned, secondary search refers to the dictionary-specific search input field which is available in entry display views. It can be used

for secondary searches as such (searches that arise as a consequence of the primary search), or as convenient input fields for starting a new primary search. In either case, the secondary search has the same type of autocomplete functionality as does the primary search input field on the home page.

For instance, while being in the entry view for *басма*, the user can initiate a new search, and click on the new desired lemma from the dropdown menu, for instance, *бајалица*. The clicked-on entry will then *replace* the entry in the current view.

Српски рјечник

баја|

Бăја +

бáја +

бăјалица +

бăјало +

бăјам +

бăјат, бăјатан +

бăјати +

Бăјац +

бăјач +

бăјанье +

Бăјко +

Бăјнă Лука +

баскйјати

10 1

Речник косовско-метохиског дијалекта

бăсма, е ж. циц. У В. 2. У В. 2 бăсма ж. 1) без оног др
 њега наведено. У RJA. 2. бăсма ж. 1).

баскйн

10 1

Illustration 5: Local search within LSG L

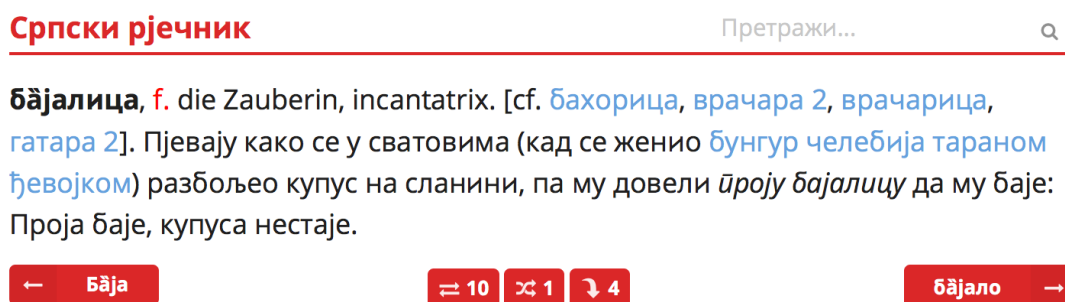


Illustration 6: Entry view replacement from secondary search

6.3.3.1 Cross-reference pop-ups

Cross-references are hyperlinks which take the user to the linked entry. But each cross-reference also functions as a trigger for a pop-up window when the user moves his or her mouse over the cross-reference. In such cases the entire cross-referenced entry is loaded in the pop-up window for quick consultation. The user can choose to follow the original cross-reference, or even some of the cross-references inside the pop-up window.

There are numerous situations in which inspecting cross-referenced entries in pop-up windows can come in handy: if the user wants to evaluate the cross-reference to determine whether it is worth opening it in a separate entry view; if the user wants to compare the current entry with its cross-reference without actually departing from the source entry, etc.

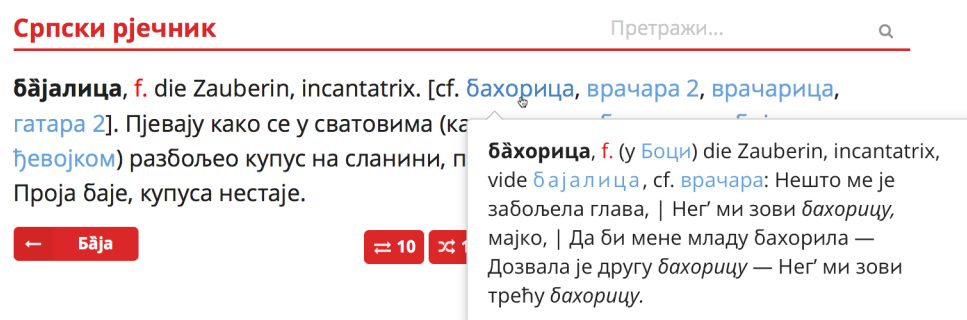


Illustration 7: Cross-reference popups

The same type of cross-reference pop-ups is available in Browsing List Views and Browsing Full-Entry Views.

6.3.3.2 *Entry collections*

I have already discussed in Chapter 1, how difficult it is to compare dictionary entries which do not appear close to one another in the dictionary macrostructure. This is fact which is well-known to dictionary users everywhere all over the world. Comparing two entries in the same print dictionary is possible if we constantly switch back and forth between two physical spaces in the macrostructure; comparing more than two entries is even more involved and can be achieved either by bookmarking the corresponding pages, or by taking extensive notes on each of the entries to be compared. Print dictionaries, because they are bound by the physical limitations of their book, cannot afford *parallel displays* of remote textual segments. In the digital medium, parallel displays *should* be possible but I know of no dictionary that has implemented this function. Some online dictionaries such as Wordnik give their users the option of collecting lists of words⁷⁸ but these consist only of links to full-entries without the possibility of displaying listed words in parallel.

Yet the need to consult more than one entry in a dictionary and compare them is a very basic one: users often follow a cross-reference, only to realize that they need to get back to the target entry, whether to compare definitions or to seek a different cross-referenced synonym. Or, users notice a word in the definition of the source entry that they need to look up — if they don't know the meaning of it, or if it's somehow related to their primary search and they think that it may contribute

⁷⁸ <https://www.wordnik.com/words/adumbrate#lists>

to the task at hand. The same applies to the need to compare entries in different dictionaries. Even though comparison in general is a basic tool of the trade in the humanities (and as such recognized as a “scholarly primitive” regardless of the discipline, see Unsworth 2005), the need for parallel displays of lexical entries has so far gone unnoticed in lexicographic literature.

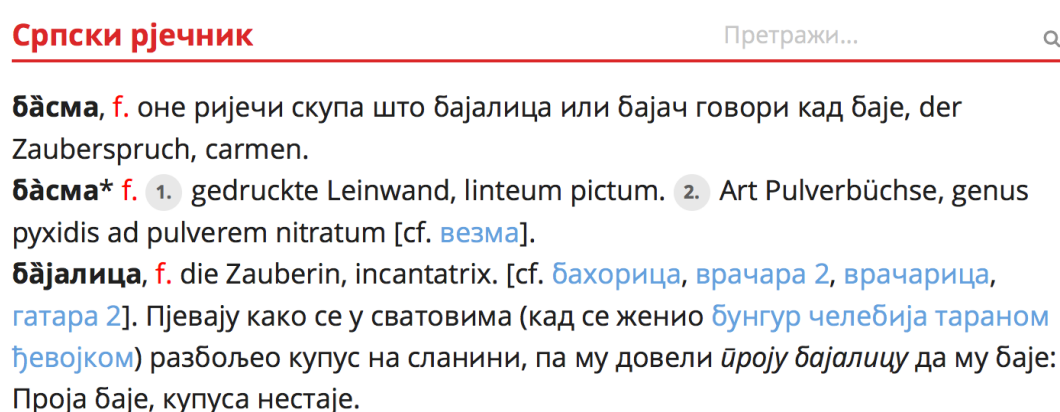
On *raskovnik.org*, as we have already seen in Illustration 5 on p. 230, the secondary search input field can be used to *replace* the contents of the current Entry View. But that is not its only function. Each lemma in the autocomplete dropdown has a plus-icon next to it. Going over the plus sign with one’s mouse reveals a pop-up which explains its functionality (“Add to this view”).



Illustration 8: Option “Add to this view” from local search

Clicking on the plus sign next to a lemma in the dropdown will *add* the corresponding entry to the current Entry View instead of replacing it. Because this functionality uses the so-called AJAX technology (asynchronous JavaScript and XML), whereby the client (browser) contacts the server (in our case the Raskovnik API) to retrieve resources without the need to reload the existing web page, the process is quick and seamless. For instance, *adding* the entry *ба̋јалица* to the above view (*ба̋сма*) will result in a collection of entries (*ба̋сма, ба̋јалица*). Each subsequent addition to the view will be displayed below the current view (see Illustration 9 on p. 234).

There is no limit to the number of entries that can be collected in this way. The add-to-view functionality exists for each of the dictionaries on raskovnik.org, which means that one can create collections of entries in each of the available dictionaries and compare multiple entries across multiple dictionaries at the same time.



Српски рјечник Претражи... 

бѣсма, f. оне ријечи скупа што бајалица или бајач говори кад баје, der Zauberspruch, carmen.

бѣсма* f. 1. gedruckte Leinwand, linteum pictum. 2. Art Pulverbüchse, genus pyxididis ad pulverem nitratum [cf. [везма](#)].

бѣјалица, f. die Zauberin, incantatrix. [cf. [бахорица](#), [врача́ра 2](#), [врачарица](#), [гатара 2](#)]. Пјевају како се у сватовима (кад се женио [бунгур челебија тараном ђевојком](#)) разбољео купус на сланини, па му довели [ђроју бајалицу](#) да му баје: Проја баје, купуса нестaje.

Illustration 9: Result of adding one entry to the current entry view

6.3.3.3 Cross-reference collections

As we have seen in the previous section, collections offer a convenient way of consolidating information from different parts of the dictionary macrostructure. Using the secondary search dropdown, however, entries can be added only one at a time.⁷⁹ It is possible, however, to add all the explicit cross-references from a currently viewed entry in one go. As can be seen above in Illustration 6 on p. 231, individual entry views come with a navigation menu below them. In addition to providing links to preceding and following entries, and contextualizing the entry

⁷⁹ One *could* implement multiple selectability in the dropdown, but we have not gone down that road yet because, on the one hand, it would make the dropdown more cluttered, and on the other, multiple selections would make sense only for words that are macrostructurally close to each other and as such appear in the same dropdown view. We may revisit this question at a later stage and, if so, with some user testing in mind.

within the dictionary macrostructure, the menu contains a button for adding all explicitly linked cross-references to the view. These entries are also retrieved asynchronously and added to the current view:

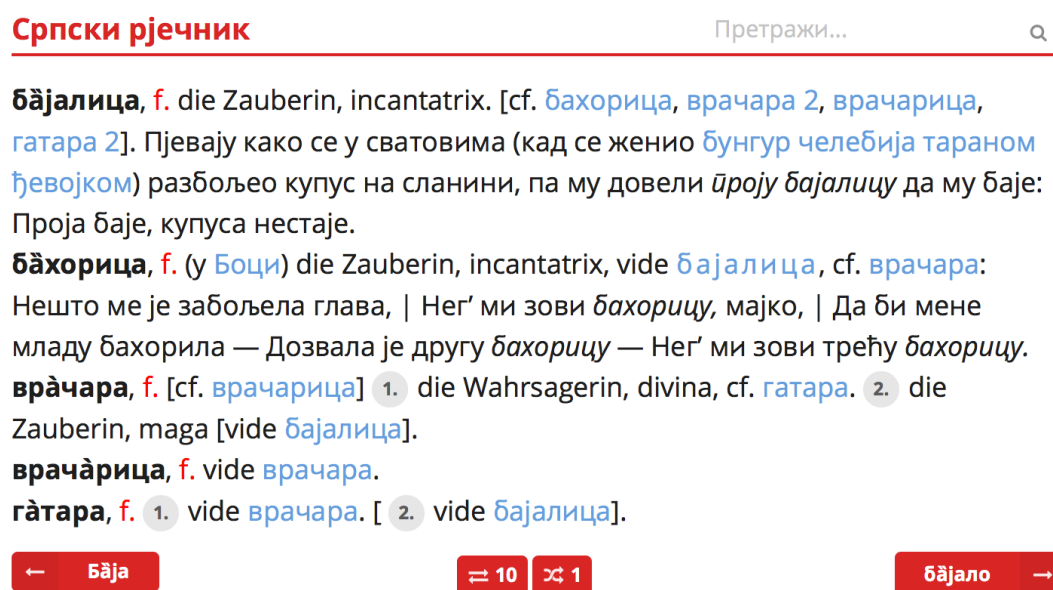


Illustration 10: Cross-reference collection view

With some entries in *LSGL* such as *сулуг*, for example, Vuk's preferred lexeme for “crazy” containing more than 20 cross-references, cross-reference collections are a useful tool for quickly comparing cross-referenced entries. It is also possible to combine the two methods: collect cross-references from an Entry View and then add additional entries to it from the secondary search dropdown.⁸⁰

6.3.4 Browsing

The digital *LSGL* affords three different types of browsing:

⁸⁰ In its current implementation, there is no way to save collected entries and collected cross-references. The user can print them out, but this is not ideal, especially, if one has to interrupt one's work and continue at a later stage. This functionality has been requested by our users and will be added in the future.

1. alphabetic (or macrostructural) browsing, which is based on the position of the entry within the dictionary macrostructure;
2. feature-based browsing, which is based on shared microstructural features among entries; and
3. geobrowsing, a subset of feature-based browsing, which is based on geolocations of the place names in dictionary entries.

6.3.4.1 *Alphabetic browsing*

Print dictionaries easily afford browsing — an “apparently aimless, but potentially pleasurable, casual reading of dictionaries” (Hartmann and James 1998: 16). Because entries in print dictionaries do not appear in isolation but are laid out with other entries on a page, browsing a print dictionary is no different from browsing any book. In most online dictionaries and dictionary portals, however, this is not the case.

The browsing function on <http://dictionary.com>, a popular online portal consists of a navigation bar at top of the browsing page containing links to all the letters of the alphabet; a middle section which contains a span of alphabetic links to the entries under the given entry; and a smaller navigation bar at the bottom of the page consisting of numbered links to the subsequent pages of the currently selected letter. The list is displayed as one long column taking up in most cases one quarter of the available screen real estate, with the other three quarters being conspicuously blank. Numbers in the page navigation bar do not indicate the total number of pages or the ranges of lemmas represented on each page. As a result, browsing on dictionary.com is largely a haphazard affair: it is practically impossible

to get a sense of one's place in the macrostructure or to jump to a particular range of entries.

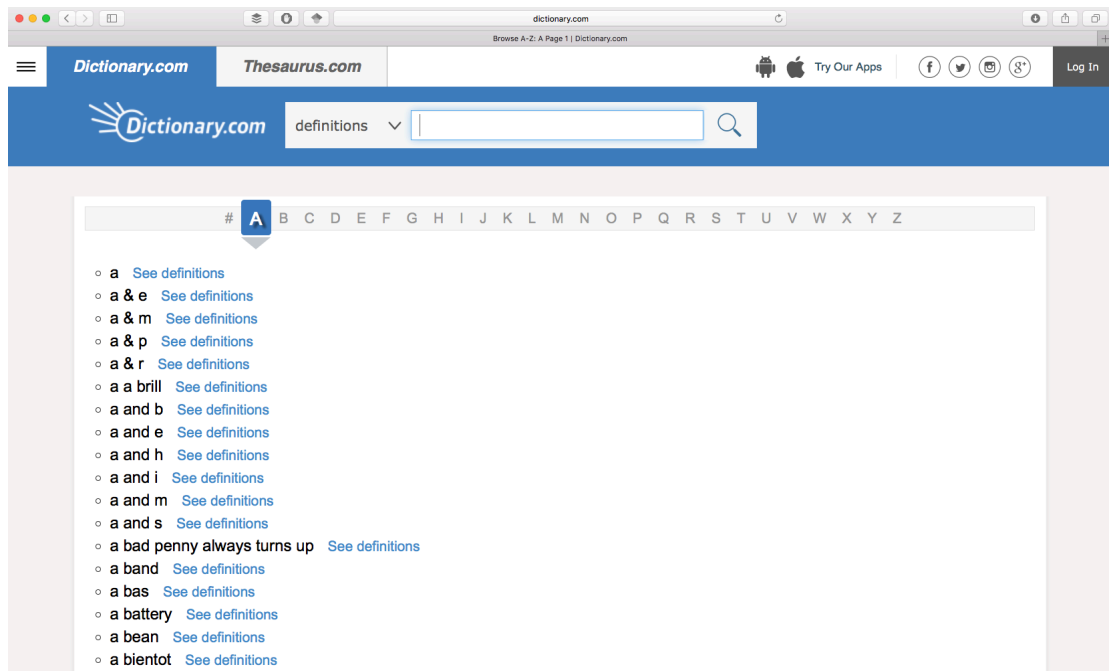


Illustration 11: The top portion of the browsing page on dictionary.com



Illustration 12: The bottom portion of the browsing page on dictionary.com

When designing the browsing interface for the digital *LSGL* and other dictionaries on *raskovnik.org*, I tried to overcome the kind of deficiencies I have described above by:

1. using both vertical and horizontal navigation bars. The vertical bar provides links to macrostructural divisions (i.e. letters of the alphabet), whereas the horizontal bar provides pagination within a selected letter of the alphabet.

2. providing a counter for the total number of items displayed per letter;
3. making sure that page numbers display the first and the last entry on the given page when the user hovers over the page number with their mouse.
4. creating not one but two browsing views: a list view, which displays lemmas only in four-columns, for easy scanning (see Illustration 13 below); and a full-entry view, which displays full entries for each page in two columns, for easy reading (Illustration 15 on p. 240).

Switching between the list view and the full-entry view is possible on each browsing page. The user's choice of the preferred view is cached.

Српски рјечник

Прикажи: ☒ списак речи ☐ целокупне одреднице				
а́рам - а́спра				
УКУПНО НА А: 582				
1	2	3	4	5
Б	а́рам	а́рати	а́риште	а́ртиџа
В	а́рамбаша	а́ратос	а́рка	а́ртовање
В	а́рамбашин	а́ратоси́љање	а́рлакача	а́ртовати
В	а́рамбашовање	а́ратоси́лати се	а́рлање	а́рфа
Г	а́рамбашовати	а́рџа	а́рлати	а́рхиђакон
Г	а́рамза́да	а́рџа	а́рли	а́рхимандрит
Д	а́рамија	а́рџлија	а́рмагани	а́рхимандритов
Д	а́рамити	а́рџлијин	а́рмија	а́рхимандритов
Ђ	а́рамљи́вање	а́рџлијински	а́рнаутин	а́рџлија
Ђ	а́рамљи́вати	а́рџлијски	а́рнаутинов	а́рџење
Е	а́рандио	а́ршљама	а́рнаутити	а́рџити
Е	а́ранђел	а́рџанас	а́рнаутити се	а́рџити се
Ж	а́ранђелов	а́рџанаса	а́рнаутка	а́рџив
Ж	а́ранђеловац	а́рџанаски	а́рнаутлук	а́ршин
Ж	а́ранђеловица	а́рџанија	а́рнаутовац	а́ршљама
З	а́ранђеловштак	а́рџија	а́рнаутски	а́ршов
З	а́ранђе́о	а́рџун	а́рнауђење	а́с
И	а́рантос	а́ргат	а́рнауџа́д	а́са ка́са лйса, а́са ка́са
И	а́рање	а́ргатџ	а́рнауче	а́слица
Ј	а́рап	а́ргатија	а́рњеви	а́сан-па́шина Па́ланка
Ј	а́рапија	а́ргатин	а́рпакáша	а́сли ва́сли
К	а́рапин	а́ргатлук	а́рпаџик	а́сна
К	а́рапинов	а́ргатовање	а́рса	а́снити
К	а́рапйиче	а́ргатовати	а́рсен	а́снити се
Л	а́рапка	а́рдџ	а́рсенија	а́сно́вит
Л	а́рапкиња	а́рдџић	а́рсице	а́спа
Љ	а́рапов	а́рџа	а́рсица	а́спида
Љ	а́рапски	а́рџа	а́рслан	а́спидин
М	а́рапчад	а́рџа	а́рсо	а́спра
М	а́рапче	а́рџа	а́ртија	
Н	а́ра́р	а́рџијетина	а́ртијетина	
Н	а́ра́рина	а́рџијешина	а́ртијешина	

Illustration 13: Browsing (list view)

All the lemmas in the list view have the quick-view feature, so that individual entries can be previewed and read without switching to the full-entry view:

☐ целокупне одреднице УКУПНО НА А: **582**

2	3	4	5
әрати арәтос аратосіѡәне аратосіѡати се әрәч әрач арәчлија арәчлијин арәчлијинскй арашләма Арбәнас Арбәнаса арбәнаскй Арбәнија әрбија әрбун әргат әргатәр аргәтија әргатин аргәтлук	ариштәнац әриште әрка арләкача әрләне әрлати әрли армәгани әрчити Арнаутка Арнаутлук арнаутовац арнаутскй арнауһење Арнаучад Арнауче әрњеви арпакаша	артијетина артијешина әртијца әртовәне әртовати әрфа архиђакон архимандрит әрчитити әрчити се әрчљив әршйн аршләма әршов Әс әса кәса лйса, әса кәса лйсице Әсан-пәшина Пәләнка	

армәгани* **m. pl.** у пјесмама vide **дар [1]**: Љубио сам у мајке једину,
 | Често сам јој слао *армајане*: | Двије дибе, четири кадифе, | И два
 ћурка куном постављена.

Illustration 14: Quick view from the browsing list view

Different numbers of items in the browsing view (582 items on page 4 of letter *a*) and the full-entry view (554 on page 4 of letter *a*) are due to the fact that the number of items in the list view refers to the number of lemmas on a given page; and in the entry view, the number of entries on a given page. Because entries can and sometimes do contain multiple lemma forms, the two numbers don't always match.

Српски рјечник

Прикажи: ☐ списак речи ☒ целокупне одреднице

УКУПНО НА А: 554

	1	2	3	4	5
А					
Б					
В					
Г					
Д					
Ђ					
Е					
Ж					
З					
И					
Ј					

Арапјинче, чета, **n.** у пјесми мјесто **Арапче**: Ал' процвили **Арапјинче** черно.
Арапка, **f.** 1. die Mohrin, Maura. [cf. **Арапкиња**]. 2. [**арайка**] eine Art harter, schwärzlicher Birnen, piri genus [pirus communis L. var.].
Арапкиња, **f.** vide **Арапка** 1.
Арапов, **a, o, adj.** vide **Арапинов**.
арапскѝ, **kā, kō, adj.** mohrisch, mauricus.
Арапчѝд, **f. (coll.)** Mohrenkinder, soboles maura.
Арапче, чета, **n.** ein Mohrenkind, puer maurus. [cf. **Арапјинче**].
арѝр[*], **арѝра**, **m.** vide **харар**.
арѝрина, **f.** vide **харарина**.
арѝти, **rām, v. impf.** plündern, expilo.
арѝтос [арѝтос] (ἀράτος) verflucht, hof's der Teufel, in malat gem! **Араѝтос** га било! **Араѝтос** ти кирије, сиђи ми с кола. Аратос ти **поскура**, не прди ми по цркви. [cf. 2 **куртала** 1].
аратосѝлање [ратосѝлање] **n. verbal. v.** **аратосѝлати се**.
аратосѝлати се, **аратѝсѝлам се** [ратосѝлати се] **v. r. impf.** кога или чега, etwas zum Teufel wünschen, detestor aliquid, exsecror.

Арѝауче, чета, **n.** ein Albaneser-Kind, puer albanus.
арѝеви, **арѝевѝ**, **m. pl.** (у **војѝ**) das Dach eines (sonst offenen) Bauernwagens, von Stäben und Rohrmatten darüber, tectum currus rustici.
арѝакаша, **f.** јечмена каша, die Gerstengrösse, ptisana. cf. [1] каша [1].
арѝаѝик* **m.** eine Art Zwiebel zum Versetzen, ceparum genus [лук са ситним лучицама, које изникну из сјемена црнога лука (allium sera L.), кад се оно посије; cf. **сијанаѝ**?].
Арса, **m.** (ист.) vide **Арсо**.
Арсен, **Арсенија**, **Арсеније**, **m.** Arsenius, Arsenius.
Арсица, **m.** dim. v. **Арса**.
арслѝн*, **арслѝна**, **m.** vide **лав**: Па ти крени звијера **арслѝна**.
Арсо, **m.** (јуж.) hyp. v. **Арсеније**. [cf. **Арса**].
арѝтија, **f.** vide **хартѝја**.
арѝјѝтина, **f.** vide **хартѝјѝтина**.
арѝјѝшина, **f.** vide **хартѝјѝшина**.
арѝјѝѝца, **f.** vide **хартѝјѝѝца**.
арѝовѝање, **n.** Berathen, Hofmeistern, consiliatio.
арѝовѝати, **тујѝм, v. impf.** kora, berathen, hofmeistern, consilior. [vide **свѝетовѝати**].

Illustration 15: Browsing (full-entry view)

All the explicit cross-references in the full-entry browsing view have the quick-view popups enabled by default.

6.3.4.2 Feature-based browsing

I define the concept of *feature-based browsing* a browsing mechanism based on collecting and displaying entries that share a certain lexicographically identifiable and computationally addressable feature (“dictionary slices” — see p. 224). I have not encountered any discussion of this type of browsing in the scholarly literature, let alone actual implementations of it. Yet I would claim that this type of browsing, which is possible only in the digital environment, can play an essential role in transforming digital editions of historical dictionaries into exploratory tools.

There are two types of dictionary slices in *LSGL*: (1) attribute-based; and (2) text-based slices. The former are created by making API calls to retrieve entries that share a particular attribute value at any level of the microstructural hierarchy. The

latter are generating by retrieving entries that have a matching string inside an explicitly encoded element at some level of the entry microstructure. The functionalities of the user interface for dictionary slices are not much different from the one we designed for macrostructural browsing: slices are paginated, can be viewed both as lists and full-entries with quick-view popups etc.

Од абе до шућура

Речи турског порекла у Вуковом *Српском рјечнику*.

Прикажи: ☐ списак речи ☒ целокупне одреднице

УКУПНО: 1850

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

рз* *m.* vide **образ 2**: *рз* и поштење.

рс[*] *m.* (у **Боци**) die Stärke eines Mannes, robur: чудна рса људскога! Ће је твоја крута снага? крути синко! | *Рсом* срзни, руком врзи, љуцки *рсу*!

рсуз* и **рсузин**, *m.* der Dieb, fur. vide **лупеж**, cf. **крадљивац**.

рсузлук* *m.* der Diebstahl, furtum. cf. [vide] **крађа**.

рувет*, *m.* vide **руфет**.

рувѐтлија[*] *m.* vide **руфетлија**.

русвâј*, русвâја, *m.* Art чудо, rei mirae genus.

руфет [**рувет**] * и **руфѐтлија** (cf. **руветлија**) [*], der Zunstgenoß, socius collegii: И терзије своје *рувѐтлије*.

сáхија [**саија**], * *m.* der Fußbote, tabellarius, cf. [vide] **књигоноша**: А *сахију* посла у Косово.

сăч*, сăча, *m.* 1. der metallene Deckel, worin z. B. die пита gebacken wird, operculum ferreum. [vide 1 **пека** 1]. 2. (у **Сријему**) од земље запушач на пећ.

сăчма*, *f.* (pl. gen. сачámâ) 1. der Schrot, grando plumbea, [cf. **менеџал**, 2 **сита** 3]. 2. сије сачмом (н. п. кукурузе), wie Schrot, während andere ökonomischer damit umgehen.

сăџăк*, *m.* Dreifuß, tripus.

сăџбаг*, *m.* vide **уплетњак**: Задје јој се *сăџбак* за мушебак.

Illustration 16: Turkisms in LSG (list-view)

Typical *attribute-based* slices in the digital *LSGL* are, for example, the slice listing Serbian words of Turkish origin⁸¹ or one which lists Greek words in the dictionary.⁸² Both are generated by similar API calls,⁸³ but the results are transformed using different XSLT stylesheets. The Turkisms list view displays Serbian lemmas, whereas the list view for Greek words displays Greek words themselves. This is for two reasons:

1. Turkisms in the print version of the *LSGL* are marked with asterisks and in the digital *LSGL* encoded as etymologies with appropriate language tags, but without actual Turkish etymons (see Section 5.3.3.4, pp. 162-163). Listing Serbian lemmas makes sense because all the words that appear in this list are, indeed, etymologically speaking, Turkisms.
2. The situation with Greek words is slightly different. There are 64 Greek words in the dictionary not counting auxiliary Greek words in Latin translations (which have a grammatical and not a semantic function, see Section 5.3.6.2, pp. 175-179). Some of those 64 words are direct etymons to the Serbian lemmas in the given entries (for instance, талас < θάλασσα, or титор < κτίτωρ), but others are mentioned only because the Serbian words are calques of the mentioned Greek words (i.e. compound Slavic words built as semantic parallels to Greek compounds, such as Богородица < θεοτόκος [she who gives birth to God], догоносни < θεοφόρος, etc.)

⁸¹ <http://raskovnik.org/isecak/turcizmi/VSK.SR>

⁸² http://raskovnik.org/isecak/grcko_nasledje/VSK.SR

⁸³ GET [http://raskovnik.org/api/v1/slice?query=/entryFree\[descendant::etym/lang\[@value="tr"\]\]/form\[@type="lemma"\]](http://raskovnik.org/api/v1/slice?query=/entryFree[descendant::etym/lang[@value=) vs. GET [http://raskovnik.org/api/v1/slice?query=/entryFree\[descendant::mentioned\[@xml:lang="el"\]\]](http://raskovnik.org/api/v1/slice?query=/entryFree[descendant::mentioned[@xml:lang=) '

In either case, quick-view popups in the list view will preview the full-entry.

Грчко наслеђе

У *Српском рјечнику* грчке речи се појављују у тумачењима порекла речи и као појашњења неких превода на латински.

Прикажи: ☒ списак речи ☐ целокупне одреднице

УКУПНО: 64

ἀλιεύς	ζουπᾶνος	κλεῖς	πολυέλαιον
ἀράτος	ἡγουμένη	κομνηνός	προκοπή
ἀργάτης	ἡγούμενος	κράβατος	προσφορά
βῶλος	ἡγούμενος	κτίτωρ	ρίπης
Δέσποινα	θάλασσα	λειτουργούντος	σαράντα
δεσπότης	τᾱλᾱς, m. (θάλασσα) die Welle, Woge, fluctus, cf. [vide] вал: Погнаше се по небу облаци, А по земљи проклети вјетрови, Подигоше на мору <i>ѡαλασε</i>.		
δεύτερος αὐτός	θρόνος	παναγία	το ρασον
δίσκος	θυμίαμα	πανήγυρις	τριανταφυλλον
διφθέρα	κάμηλος	παράκλησις	τριημερίζω
δραγμαευ	κασσίτερος	παρουσίαν	υπέρπυρον
δρόμος	κελλίον	πάτος	χωριάστες
Εἰρήνη	κεραμῖς	περιτραχήλιον	χωριάτης
εἰχών	κιβωτός	ποίησις	ῶρα
εὐθυνότης			
ἐυσεβής			
εὐφημιστικῶς			

Illustration 17: List view of Greek words in *LSGL*

A deeply encoded dictionary such as *LSGL*, in combination with a flexible API that takes Xpath expressions, can produce a large number of dictionary slices, each of which can be explored in a browsing interface. It would be easy to imagine a data-enrichment workflow in which scholars or PhD students further annotate the XML version of *LSGL* for their own needs, but with an understanding that their annotations can be used as a basis for the creation of publicly accessible dictionary slices.

Грчко наслеђе

У *Српском рјечнику* грчке речи се појављују у тумачењима порекла речи и као појашњења неких превода на латински.

Прикажи: ☐ списак речи ☒ целокупне одреднице

УКУПНО: 64

àlās, алáса, **m.** (ἀλιεύς, der Fischer)

vide **рибар**.

арàтос [**арантос**] (ἀράτος) verflucht, hol's der Teufel, in malam rem! *Αραῖος* га било! *Αραῖος* ти кирије, сиђи ми с кола. Аратос ти **поскура**, не прди ми по цркви. [cf. 2 куртала 1].

àргатин, **m.** (ἀργάτης), mercenarius. cf. [vide] **надничар**.

благочàстивй Рйм (nach εὐσεβής) fromm, rechtgläubig, Gott recht verehrend, pius, orthodoxus. Срби кажу: први је Рим био *блаїочасїиви*, па ће бити и пошљедњи.

богòносац, сца, **m.** (nach dem griech. θεοφόρος) н. п. Игњат богоносац, Ignatius theophorus (den 20. Dezemb.

stannum. [cf. **калај**, **цин**].

крèвет, **m.** 1. das Bett, lectulus (grabatus, κράβατος). [vide **постеља**]. cf. **одар**. 2. око варнице поузвишено мјесто, на које чаругџије коже из варнице међу.

мèтанија, **f.** (μετάνοια) die tiefe Verneigung und Bekreuzigung, prostratio, adoratio: ударити десет, или сто метанија. Метаније су мале и велике: мале: кад се човјек прегне и дохвати рукама до земље, а велике: кад клекне на кољена и на руке, па дохвати главом до земље.

мйнеј, **m.** das Menäon (Μηνᾶιον), ein Kirchenbuch, menaeum.

нàруквица, **f.** 1. das Armband,

Illustration 18: Greek words (full-entry view)

A typical example of the *text-based* (as opposed to *attribute-based*) slice would be the one entitled *Изгубљени у преводу* [Lost in translation], which lists 941 entries (and 977 lemma forms) from *LSGL* that have imprecise translation equivalents in German or Latin in at least one of the senses (for instance, *трѣскавица* → Art Tanzes [kind of dance] or *ба̀бини зѹби* → herbae genus [kind of plant]).

Изгубљени у преводу

Лексеме из Вуковог *Српског рјечника* са непрецизним еквивалентима на немачком или латинском, нпр. *Шрѣскавица* → нем. Art Tanzes („врста игре“) или *бабини зѹби* → лат. herbae genus („врста биљке“).

Прикажи: ☐ списак речи ☒ целокупне одреднице

УКУПНО: 941

1	2	3	4	5	6	7	8

авáјлија* **f.** (у **Уж. н.**) некака јабука, Art [Winter-]Apfel, mali genus [pirus malus L. var.].

авгута́р, авгута́ра, **m.** сух рибли [суха рибли] мријест, који [која] се (понајвише у **Скадру**) вади из рибе скакавице, Art Pökelrogen, ova piscium condita. Августар се једе и уз часни пост сваки дан, као и хајвар.

а́здија* **f.** (ст.) Art eines langen Oberkleides, togae genus: На њему је коласта *аздија* — У скут свилен коласте *аздије*.

а́лет* **m.** eine Art Kleid, vestis genus: Ен' оно је Алил челебија! | У *алеџу* у

блор [главор]

блòруша, **f.** (у **Ц. г.**) некака змија, Art Schlange, serpentis genus. Она је дуља и тања од блаора [cf. **блаворуша**].

бљушт, **m.** некака трава, која горе расте као лоза, а коријен јој је као ротква, Art Gewächs, herbae genus [tamus communis L.].

бòбица, **f.** (у **Дубр.**) ситни боб, Art Bohne, fabae genus [vicia faba L. var. equina].

бòбња, **f.** (у Биогр.) као зелена паприка, и кува се за јело, Art Gemüse, oleris genus [зелен плод од hibiscus esculentus L.].

Illustration 19: “Lost in translation” (full-entry view, excerpt)

A large portion of hyperonymic translations in *LSGL* are used in entries describing flora and fauna, but there are also examples of specific types of weapons (e.g. *кавалија ѿоѹ*), clothing (*каваѡг*), games (*ѿз кола кѹчкѧ*), illnesses (*ѹлѧзина*) etc. Many of these words will be completely unfamiliar to speakers of modern Serbian.

6.3.4.3 Geobrowsing

Geobrowsing is a type of feature-based browsing which takes as a point of departure not words but geolocations displayed on a map. In Section 5.3.5.2, we already discussed the use of `<rs>` (referring string) inside `<usg type="diatopic">` for marking up the text string of the place name in combination with the attribute key

holding the id of a corresponding authority record as its value. The authority record for each place contains, in addition to the name of the place, some supplemental information about the place, including latitude and longitude. This encoding makes it relatively easy to collect all the place names from diatopic usage labels and display them as “pins” on a zoomable map.

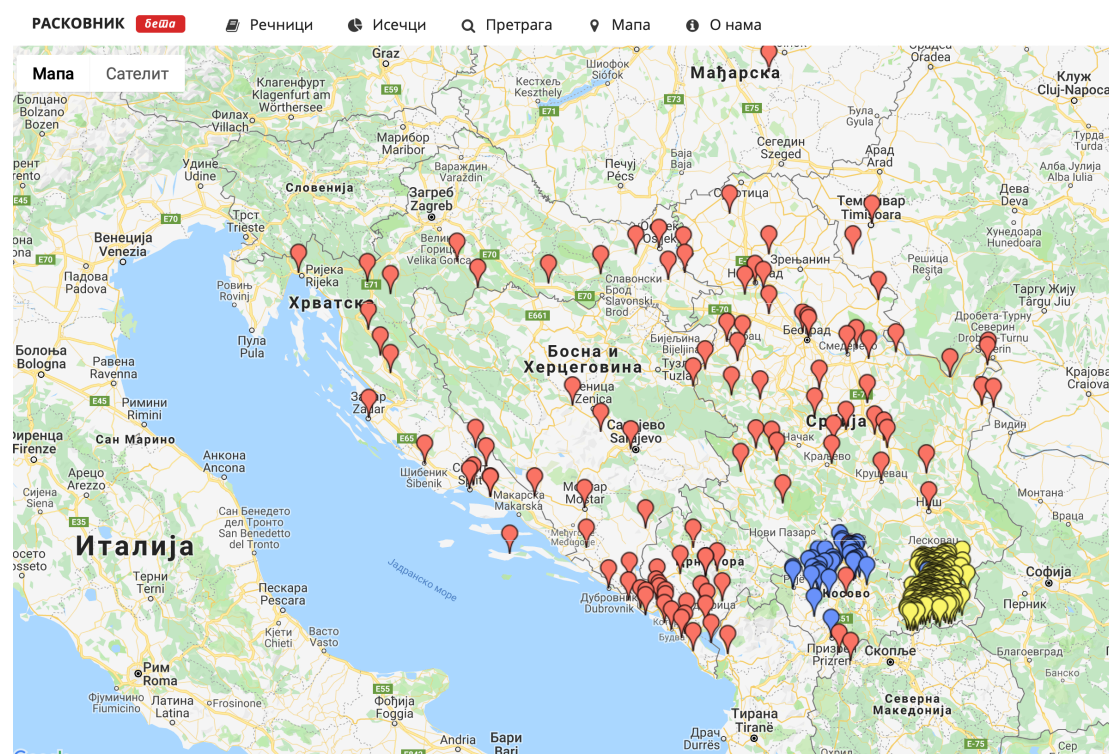


Illustration 20: Raskovnik Geobrowser (geolocated place names in LSG are marked in red)

Illustration 20 above shows a geospatial interface to three different dictionaries: the LSG (in red), the *Dictionary of the Kosovo-Metohija Dialect* by Gliša Elezović (Елезовић 1932, 1935) in blue, and the *Dictionary of South Serbian Dialects* by Momčilo Zlatanović (Златановић 1997) in yellow. The distribution of geolocated entries on the map echoes the essential difference between Vuk’s dictionary and the later dialect dictionaries. The concentration of geolocated entries in Elezović and Zlatanović clearly visualizes their focus on a single dialect, in one particular region;

LSGL, on the other hand, contains diatopic usage labels from a much wider geographic area, covering practically the entire Serbo-Croatian diasystem.

Clicking on a marker on the map displays a pop-up window with the name of the place. Clicking on the place name generates the same kind of browsing interface, which is used for dictionary slices: a collection of entries (in list-view or entry-view) that are associated with a particular place (see Illustration 21). This way, the user can use the map as a separate access-point for exploring the dialectal vocabulary in the dictionary.

Бока Которска географска област



Прикажи: ☒ списак речи ☐ целокупне одреднице

УКУПНО: 673

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

а̋к	ба̋тун	блѣска	брѣме
а̋мат	ба̋хорити	блѣснути	бр̋кљача
а̋ца̋л	ба̋хорица	боб̋отати	бр̋ница
ба̋бовић	бе̋дрени̋пра̋г	бог̋ата̋ш	бр̋хан
ба̋бушка	бе̋рба	бог̋ора̋дни̋к	бр̋цнути
ба̋да̋л̋	бе̋сти̋ј̋нски̋	Б̋джић	бр̋ч
ба̋де̋л̋	би̋љешка	боке̋љски	бр̋ча
ба̋к	б̋јегӳница	б̋о̋к̋ин	бр̋чни̋
ба̋ка̋ла̋ј	б̋јегӳнче	бра̋танић	Бӳгар
ба̋нди̋жа̋н	бл̋а̋ти̋ште	бра̋таница	бӳран
ба̋н̋ђен	бл̋е̋ка	бра̋че	Ва̋ра
ба̋ња	бл̋е̋ка	бр̋дун	ва̋рица

Illustration 21: Raskovnik “geoslice” for the region of Boka Kotorska

At the moment, each place on a map is indicated by a pin regardless of the number of entries that are associated with a given location. Future work may include the development of markers of different sizes, or marker clusters, which would indicate the relative size of entries associated with a particular geolocation. Another possible area of improvement would be the development of so-called shape files for regions in order to distinguish them from cities or towns.

6.3.5 *Serendipitous discovery*

In Chapter 1, I described three types of serendipitous discovery afforded by the dictionary: by proximity, by relation and by incidence (see p. 77). The user interface of the digital *LSGL* facilitates all three types of accidental discovery. Some of the functionalities I have described above already contribute to serendipitous discovery: the quick-view mechanism for cross-references, for instance, encourages discovery by relation. The primary and secondary search dropdowns can display lemmas that the user was not actively seeking for, but that appear in the proximity of the sought headword. There are also additional mechanism built into the user interface.

The single-entry view automatically displays a navigation bar at the bottom of the entry, with links to the previous and the following entry (see Illustration 6 on p. 231). Clicking on either of those entry links (which appear not as anonymous links to “next” and “previous” entry, but are fully named instead) lets the corresponding entry “slide in” asynchronously, i.e. without having to load a new page. The newly loaded entry will itself display links to the macrostructurally closest entries, so turning the single-entry view into a kind of window onto a seemingly infinite slider of entries.

It is also possible, using the same navigation bar, to contextualize the given

entry by collecting and displaying it together with 10 entries that precede it and 10 entries that follow it:

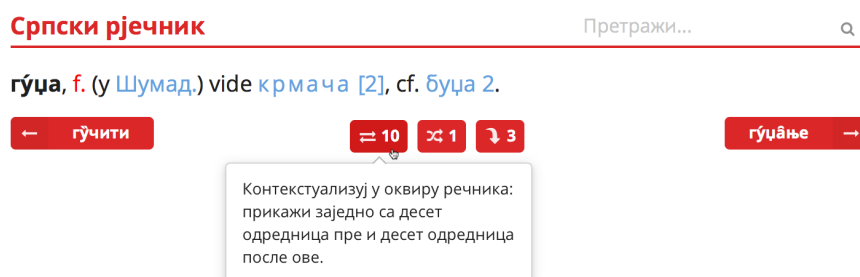


Illustration 22: Contextualize-button.

The resulting view is similar to browsing view, except the entries are displayed in one instead of two columns:



Illustration 23: Contextualized view of a single entry

The above discussion has shown how an innovative, exploratory interface of a historical dictionary can be built on top of a flexible API and rich, semantic encoding of the dictionary text. The implications of this approach go beyond editions of historical dictionaries. APIs can not only function as delivery mechanisms for text as a service, but can help with the consistent detachment of form and content in digital interfaces. Most importantly, however, this chapter has shown how a deep understanding of information-related behaviours that are *afforded* by the dictionary can be used in the digital medium to both reflect some important though sometimes overlooked aspects of the paper-based source (alphabetic browsing, contextualized reading, serendipitous discovery etc.); and to create exploratory, interactive views that are only possible in a digital edition (entry collections, feature-based browsing, dictionary slices, geobrowsing etc.).

CONCLUSION

In my most cynical moments, I tend to think of digital humanities as the ultimate revenge of global capitalism on the traditional humanities. There is a troubling reliance on automation and cultural super-efficiency as a mode of intellectual activity within DH: we try to process as much text as possible in as little time as possible. A lot of what we do — statistical analysis of big data, distant reading, data visualization, geospatial mapping etc. — is quite problematic because it can and it often does take texts out of their contexts.

When Derrida famously proclaimed that there is nothing outside the text (Derrida 1997: 158), he was not suggesting that the world of literary communication is hermetic, claustrophobic and strictly linguistic. Rather, he was emphasizing the essential role of context, that is, of extra-textual conditions, in the construction and creation of a text. While context is always already present (at the moment of text production), it is also already lost (at the moment of reception). Every sign, every inscription — indeed every text — exists in a Derridean mode of survival: it can be deciphered, read and analyzed in the radical absence of its creator. Texts outlive

their authors and the historical contexts in which they were written. This is something that we as readers, editors and writers have to live (and die) with.

Texts, however, are never transmitted entirely without at least some traces of their context(s). The dynamic relationship between text and context is reflected, for instance, in the interconnection of a text and its paratexts — that heterogeneous cluster of discursive elements that accompany a text, such as title pages, tables of contents, dedications, prefaces, illustrations, annotations (Genette 1997). Some of the paratextual features, such as title pages, provide clues about the chain of production (publisher, series, date of publication, editor); dedications and acknowledgements contextualize the author's social network; annotations provide supplementary information that the editor believes a reader will need or find useful, etc. This, I would claim, applies equally to the traditional, print annotation mechanisms such as footnotes or endnotes *and* to textual markup, even though the latter may remain hidden from end-users in the GUI.

Even though annotations and textual markup may at first appear to function as external, peripheral additions to the “core” text, they are, in fact, neither auxiliary nor accidental. The paratextual sphere is not just integral in shaping the text's semantic field, it is also, as Genette has argued, a privileged place of influence and control, “a fringe of the printed text which in reality controls one's whole reading of the text” (Genette, 1997: 2). In the context of the digital *LSGL*, I have shown, how TEI markup stands at the intersection of the text's linguistic properties (as a model-*of*) and its potential functionalities in the digital environment (as a model-*for*). A dictionary “slice”, for instance, is both a feature of the text and a precondition for a type of functionality of a digital edition.

Yet if, in a nightmarish, apocalyptic endgame of one particular trend within Digital Humanities, all texts, which are by definition complex, idiosyncratic and – unless we’re dealing with supermarket receipts – potentially ambiguous, are turned into flat, unstructured *data*, we run the risk of Digital Humanities becoming the Grand Decontextualizer of the traditional humanities. Big data, distant reading and algorithmic criticism present only one possible dimension of scholarly engagement with textuality. And it is an important one. Not just because it reminds us of our all too human limitations (how much can each of us actually read in a lifetime? and how do we come to terms with the fact that no matter how much we read, we are always reading but a fraction of what has been written), but because patterns that can be established, analyzed and visualized using various DH methods can tell us an important story which is broader in terms of scope and less impressionistic in terms of method than many pronouncements and conclusions which are deduced from the close reading of a small number of texts. But this data-centric story is also always partial because the same word or the same pattern in two different texts can play very different roles. We can study, for instance, word distribution patterns in British 19th-century novels, and we can argue that the trends we observe in this corpus are representative, both of the novel as such, and of British novels in a particular epoch, but we cannot help but notice that distant reading methods obscure the fact that not all texts were born equal. I am not even considering here conservative frames of reference such as literary or poetic quality. But I *am* talking about impact: some texts get purchased more than others; some have higher circulation in libraries; some are discussed in the press at great length, while others remain hidden from the general public. Distant reading techniques reveal patterns that are not identifiable with traditional reading techniques, but they also reveal

patterns that can be misleading or skewed by the very fact that they are based on the radical equality of treating all texts as unweighted, context-free data. Cultural production and cultural consumption are not and have never been completely egalitarian.

In my work on the digital edition of Vuk Stefanović Karadžić's *Lexicon Serbico-Germanico-Latinum*, I have engaged in a different kind of digital humanities: one based on small data and not-too-distant reading methods. I have analyzed my object of study — a historical dictionary — as a complex socio-cultural assemblage, not just a data source. I have shown how a thorough analysis of the dictionary's functional affordances can inform the way we model and encode legacy content and the way we conceptualize and implement reading interfaces. Finally, I have shown how this type of transdisciplinary digital humanities, which is guided not simply by the speed and the scope of text processing, but by the intensity of engagement with the remediation of a cultural artifact, can serve as an important corrective to some eccentric scholarly enterprises, such as the efficiency-obsessed Function Theory of Lexicography or a tunnel-visioned attack on the Text Encoding Initiative by those who believe that creating functional, interactive editions is only possible by using ad-hoc encoding vocabularies and mixing them with engineering classes.

A full articulation of a non-utilitarian, medium-specific humanistic theory of electronic lexicography was certainly beyond the scope of this dissertation. But I do believe that I have set the stage for one such undertaking. A medium-specific theory of electronic lexicography, situated within the wider field of digital humanities, could explore the notion of lexicography as a “trading zone” between different epistemic cultures and the dictionary as a “boundary object” (Galison 1997;

Gorman 2002; Gorman 2010) which can mean different things to different parties without invalidating their individual paradigms. If developed, it would:

1. offer a unified theory of electronic lexicography that would be equally applicable in the production of contemporary dictionaries and the retrodigitization of historical dictionaries;
2. move beyond the primitive instrumentality of the Function Theory of Lexicography and explore the scholarly use of electronic dictionaries as well as the question of how the digital medium can satisfy dictionary research interests beyond what was already possible in print;
3. contribute to the conceptualization of a new type of lexicographic research and a creation of new digital resources that will be better equipped to shed light on lexicographic developments across national and cultural borders.

A new theory of electronic lexicography and the transformation of historical lexicography into a virtual platform for historical semantics (Charlier 2005) will in my mind be possible only if digital humanities does not become a one-trick pony. For us to push the boundaries of lexicographic research *and* lexicographic production, we need to make sure that we use all the available methods. That also means that we have to insist on the continued importance of structured, contextualized data. The development of deeply encoded, mutually interoperable XML editions of historical dictionaries and the development — not just in theory but also in practice — of digital research infrastructures will play an important role in this undertaking.⁸⁴

⁸⁴ In terms of lexicographic infrastructures, previous work in this field includes a proposal for the creation of a federation of heterogeneous dictionary databases (FHDDDB) whose structure would be mapped to a generalized dictionary description language (GDDL) and queried using a generalized dictionary query language (GDQL) to facilitate the exchange of information

It has already been suggested — in somewhat Utopian terms — that the “elektronische[...] Verknüpfung aller großen historischen Wörterbüchern Europas” [electronic linking of all the great historical dictionaries of Europe] (Haß 2011: 52) could help scholars explore the transnational and multilingual nature of lexicographic discourse. The importance of this point can hardly be overestimated. The true potential of the digital medium in the study of dictionaries is not only in the access it provides to our national traditions but precisely in the possibility to study wider cultural developments beyond national borders. A truly comparative history of lexicography is yet to be written. I would claim that its success will to a large extent be shaped by the scholarly use of technology in the years to come.

between various remotely stored dictionaries in different formats and the end user (Patrick *et al.* 2000). A human-centered application of Semantic Web technologies lead one author to propose the language grid — an intercultural collaboration experiment that treats existing language services as atomic components and allows users to combine existing components into new services that would be available on the Internet (Calzolari 2008). Hayashi and Ishida (2006) focus on dictionary access within the proposed language grid and describe in greater detail a unified abstract dictionary model that can represent a range of machine-readable human dictionaries. Even though the paradigm of distributed and interoperable lexicons has widely been discussed and invoked, “little has been made for its practical realisation” (Calzolari 2008: 38). A potentially fruitful avenue of research is the application of the Linked Data paradigm (Berners-Lee 2006) to the publication and representation of linguistic resources on the Web (see Chiarcos *et al.* 2012).

APPENDICES

APPENDIX I

RASKOVNIK API

The interactive version of the API documentation is available at <http://docs.raskovnikapi.apiary.io>.

A static snapshot of the documentation can be found on the following pages.

Raskovnik API

Raskovnik API is an application programming interface which allows clients to perform context-based searches on TEI-encoded dictionaries hosted on raskovnik.org.

Allowed HTTP requests:

GET : Get a resource or list of resources

Description of Usual Server Responses:

- 200 **OK** - the request was successful.
- 204 **No Content** - the request was successful but there is no representation to return (i.e. the response is empty).
- 400 **Bad Request** - the request could not be understood or was missing required parameters.
- 401 **Unauthorized** - authentication failed or user doesn't have permissions for requested operation.
- 403 **Forbidden** - access denied.
- 404 **Not Found** - resource was not found.
- 405 **Method Not Allowed** - requested method is not supported for resource.
- 412 **Precondition Failed** - requested parameters are not valid.
- 500 **Internal Server Error** - API has been broken.

Pagination and Headers:

Raskovnik API supports various response headers and pagination for the easy retrieval of longer responses. Most of the endpoints described below have optional parameters to support pagination.

Headers

Consider the following example and description of the generated headers as follows:

If a given query returns the expected results to the exact total of **90**, and, if limit is **10**, then there would be 9 (90 / 10) pages, i.e. **total divided by limit**. Using these values, the client can retrieve various results and navigate forth and back to the next set of results.

- **HTTP Headers**
 - **X-Current-Page** - Shows the current page of the results-subset, e.g., can be 1 to 9 in the above example.
 - **X-Limit** - Shows the number of results per page, e.g., 10.
 - **X-Total-Pages** - Shows the total number of pages, e.g., 9.
 - **X-Total-Results** - Shows the total number of results, e.g., 90.

Pagination

Along with various headers, one additional link header is provided by the API to smoothly navigate to the first, previous, next and the last page. The url is automatically created by the API itself along with the set user values for "limit" and "page".

Consider the following **link** header, where the client is on Page 4:

```
<http://raskovnik.org/api/v1/slices?query=//entryFree[descendant::etym/lang[@valu
<http://raskovnik.org/api/v1/slices?query=//entryFree[descendant::etym/lang[@valu
<http://raskovnik.org/api/v1/slices?query=//entryFree[descendant::etym/lang[@valu
<http://raskovnik.org/api/v1/slices?query=//entryFree[descendant::etym/lang[@valu
```

- **Link Header**
 - **first** - Shows the url for the first page of results, marked with rel="first"
 - **prev** - Shows the url for the previous page of results, marked with rel="prev"
 - **next** - Shows the url for the next page of results, marked with rel="next"
 - **last** - Shows the url for the last page of results, marked with rel="last"

Note: Limit can be set to `0` in order to fetch all results in a single page and remove pagination link.

Resource Group

ENTRIES

Returns entries based on the lemma (dictionary headword) in the given dictionaries.

GET

[List First Lemma Entry](#)

`/v1/entries{?lemma,dict,page,limit,transform,format,group,wrap,contains}`

Example URI

GET <http://raskovnik.org/api/v1/entries?lemma=a6a&dict=&page=&limit=1&transform=&format=&group=&wrap=&contains=>

URI Parameters

[Hide](#)

lemma	<code>String</code> (required) Example: a6a a lemma to be searched
dict	<code>String</code> (optional) Default: all retrieves resources for all or specific dictionaries based on their ids. Multiple ids should be comma-separated, for instance <code>dict=VSK.SR,MZ.RGJS</code> .
page	<code>Number</code> (optional) Default: 1 page of results expected by client.
limit	<code>Number</code> (optional) Default: 10 Example: 1 maximum number of records expected by client.
transform	<code>String</code> (optional) Default: false name of the XSLT file (in the \$TRANSFORMATION directory) that should be applied to retrieved resources before returning them.
format	<code>string</code> (optional) Default: xml results expected by client should be in XML or JSON. Choices: <code>xml</code> <code>json</code>
group	<code>string</code> (optional) Default: false group results by dictionary. Choices: <code>true</code> <code>false</code>
wrap	<code>string</code> (optional) Default: true specifies whether the returned query results are to be wrapped with a surrounding <code><exist:result></code> element. Choices: <code>true</code> <code>false</code>
contains	<code>string</code> (optional) Default: false specifies whether the lemma returned should be an exact match of the query string or if it should contain it. Choices: <code>true</code> <code>false</code>

Response `200`

[Show](#)

SLICES

Returns a dictionary "slice" on the basis of the submitted XPath query in the given dictionaries. The client should be familiar with the XML structure of the dictionary.

GET

[List First 10 Slices](#)

`/v1/slices{?query,dict,page,limit,transform,format,group,wrap}`

Example URI

GET [http://raskovnik.org/api/v1/slices?query=/entryFree\[descendant::etym/lang\[@value='tr'\]\]/form\[@type='lemma'\]&dict=&page=&limit=&transform=&format=&group=&wrap=](http://raskovnik.org/api/v1/slices?query=/entryFree[descendant::etym/lang[@value='tr']]/form[@type='lemma']&dict=&page=&limit=&transform=&format=&group=&wrap=)

URI Parameters[Hide](#)

query	XPath (required) Example: <code>//entryFree[descendant::etym/lang[@value='tr']]/fc</code> an XPath expression to generate the slice. For more details on XPath, see https://www.w3schools.com/xml/xpath_intro.asp .
dict	String (optional) Default: all retrieves resources for all or specific dictionaries based on their ids. Multiple ids should be comma-separated, for instance <code>dict=VSK.SR,MZ.RGJS</code> .
page	Number (optional) Default: 1 page of results expected by client.
limit	Number (optional) Default: 10 maximum number of records expected by client.
transform	String (optional) Default: false name of the XSLT file (in the \$TRANSFORMATION directory) that should be applied to retrieved resources before returning them.
format	string (optional) Default: xml results expected by client should be in XML or JSON. Choices: xml json
group	string (optional) Default: false group results by dictionary. Choices: true false
wrap	string (optional) Default: true specifies whether the returned query results are to be wrapped with a surrounding <code><exist:result></code> element. Choices: true false

Response [Show](#)**CONTEXT**

Returns the siblings of the given entry wrapped in divs (with `@type` attribute having possible values `left`, `current` and `right` for the preceding, current and following siblings respectively.)

GET[List Context](#)

```
/v1/context{?entry,left,right,return,transform,format,group,wrap}
```

Example URI

GET <http://raskovnik.org/api/v1/context?entry=VSK.SR.nowtape&left=1&right=1&return=&transform=&format=&group=&wrap=>

URI Parameters[Hide](#)

entry	String (required) Example: <code>VSK.SR.nowtape</code> an entry whose context is requested.
left	Number (optional) Default: 10 Example: 1 count of the expected preceding-siblings for the given entry. Maximum value can be 50 .
right	Number (optional) Default: 10 Example: 1 count of the expected following-siblings for the given entry. Maximum value can be 50 .

return	string (optional) Default: both return only preceding siblings, if set to "left", following siblings, if set to "right", else both. Choices: both left right
transform	String (optional) Default: false name of the XSLT file (in the \$TRANSFORMATION directory) that should be applied to retrieved resources before returning them.
format	string (optional) Default: xml results expected by client should be in XML or JSON. Choices: xml json
group	string (optional) Default: false group results by dictionary. Choices: true false
wrap	string (optional) Default: true specifies whether the returned query results are to be wrapped into a surrounding <exist:result> element. Choices: true false

Response [Show](#)**RANGE**

Returns a range of entries given two entry ids. It includes the "from" and "to" entries along with the result.

GET [List Range](#)**Example URI**[GET http://raskovnik.org/api/v1/range?from=MZ.RGJS.a6a1&to=MZ.RGJS.a6e1e&transform=&format=&group=&wrap=](http://raskovnik.org/api/v1/range?from=MZ.RGJS.a6a1&to=MZ.RGJS.a6e1e&transform=&format=&group=&wrap=)**URI Parameters**[Hide](#)

from	String (required) Example: MZ.RGJS.a6a1 an entry where search should begin from.
to	String (required) Example: MZ.RGJS.a6e1e an entry till search should last.
transform	String (optional) Default: false name of the XSLT file (in the \$TRANSFORMATION directory) that should be applied to retrieved resources before returning them.
format	string (optional) Default: xml results expected by client should be in XML or JSON. Choices: xml json
group	string (optional) Default: false groups results by dictionary. Choices: true false
wrap	string (optional) Default: true specifies whether the returned query results are to be wrapped within a surrounding <exist:result> element. Choices: true false

Response [Show](#)

RANDOM

Returns a random entry.

GET`/v1/random{?dict,transform,format,group,wrap}`[List Random Lemma Entry](#)**Example URI**

GET <http://raskovnik.org/api/v1/random?dict=&transform=&format=&group=&wrap=>

URI Parameters[Hide](#)

dict `String` (optional) **Default:** all
pool of dictionaries to consider. Multiple ids should be comma-separated, for instance `dict=VSK.SR,MZ.RGJS`.

transform `String` (optional) **Default:** false
name of the XSLT file (in the \$TRANSFORMATION directory) that should be applied to retrieved resources before returning them.

format `string` (optional) **Default:** xml
the format expected by the client.

Choices: `xml` `json`

group `string` (optional) **Default:** false
groups results by dictionary.

Choices: `true` `false`

wrap `string` (optional) **Default:** true
specifies whether the returned query results are to be wrapped within a surrounding `<exist:result>` element.

Choices: `true` `false`

Response `200`

[Show](#)

APPENDIX II

FULL-TEXT OF TEI-ENCODED *LSGL*

The full-text of the TEI-encoded *LSGL* is available from <https://github.com/ttasovac/lsgl>.

BIBLIOGRAPHY

Adams, Douglas and John Lloyd (1984). *The Meaning of Liff*. New York: Harmony Books.

Adams, Douglas and John Lloyd (1990). *The Deeper Meaning of Liff*. New York: Harmony Books.

Agarwal, Naresh Kumar (2015) Towards a definition of serendipity in information behaviour., *Information Research: An International Electronic Journal*, 20(3): n3.

Ahlsweide, Thomas and Martha Evens (1988) Generating a Relational Lexicon from a Machine-Readable Dictionary, *International Journal of Lexicography*, 1(3): 214-37.

Ahlsweide, Thomas E (1985). A Tool Kit for Lexicon Building. *Proceedings of the 23rd annual meeting on Association for Computational Linguistics*, 268-76. Chicago, Illinois: Association for Computational Linguistics.

Aitchison, Jean (2003). *Words in the Mind: An Introduction to the Mental Lexicon*. Malden, MA: Blackwell Publishing.

- Aitken, A. J. and Paul Bratley (1967) An Archive of Older Scottish Texts for Scanning by Computer, *English Studies*, 43: 60-1.
- Aitken, A. J. (1971). Historical Dictionaries and the Computer. In Wisbey, R. A. (ed.), *The Computer in Literary and Linguistic Research. Papers From a Cambridge Symposium*, 3-17. Cambridge: Cambridge University Press.
- Albin, A. (1970) The Creation of the Slaveno-Serbski Literary Language, *The Slavonic and East European Review*, 48(113): 483-91.
- Aleksić, Tatjana (2009) National Definition through Postmodern Fragmentation: Milorad Pavić's Dictionary of the Khazars, *The Slavic and East European Journal*, 53(1): 86-104.
- Alison, James (2006) Automatism, Arbitrariness, and the Oulipian Author, *French Forum*, 31(2): 111-25.
- Alshawi, Hiyan, et al. (1989). Placing the Dictionary On-Line. In Boguraev, Bran and Ted Briscoe (eds.), *Computational Lexicography for Natural Language Processing*, 41-63. London and New York: Longman.
- Amsler, Robert A (1981). A Taxonomy for English Nouns and Verbs. *Proceedings of the 19th Annual Meeting of the Association for Computational Linguistics*, 133-8. Association for Computational Linguistics.
- Amsler, Robert Alfred (1980). The Structure of the Merriam-Webster Pocket Dictionary. University of Texas at Austin.
- Amsler, Robert Alfred (1984) Machine-Readable Dictionaries, *Annual Review of Information Science and Technology*, 19: 161-209.
- Andersen, Birger and Richard Almind (2011). The Technical Realization of Three Monofunctional Phrasal Verb Dictionaries. In Fuertes Olivera, Pedro and Henning Bergenholtz (eds.), *E-Lxicography: The Internet, Digital Initiatives and Lexicography*, 208-29. London and New York: Continuum.
- Anderson, Benedict R (1991). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London; New York: Verso.

- Arnold, Ellen (ed.) (2007). *The Salt Companion to Carter Revard*. Cambridge: Salt Publishing.
- Aspray, A, et al. (eds.) (1990). *Computing Before Computers*. Ames, Iowa: Iowa State University Press.
- Atkins, B. T. S (ed.) (1998). *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators*. Tübingen: Niemeyer.
- Atkins, Beryl T Sue (1996). Bilingual Dictionaries: Past, Present and Future. 515-46. Göteborg: Department of Swedish, Göteborg University.
- Atkins, Beryl TS and Krista Varantola (1997) Monitoring Dictionary Use, *International Journal of Lexicography*, 10(1): 1-45.
- Atkins, Sue and Michael Rundell (2008). *The Oxford Guide to Practical Lexicography*. Oxford and New York: Oxford University Press.
- Auty, Robert (1958) The Linguistic Revival Among the Slavs of the Austrian Empire, 1780-1850: The Role of Individuals in the Codification and Acceptance of New Literary Languages, *The Modern Language Review*, 392-404.
- Avram, Henriette D., et al. (1965). *A Proposed Format for a Standardized Machine-Readable Catalog Record; A Preliminary Draft*. Washington: Library of Congress.
- Badenoch, Alexander and Andreas Fickers (2010). Europe Materializing? Toward a Transnational History of European Infrastructures. In Badenoch, Alexander and Andreas Fickers (eds.), *Materializing Europe: Transnational Infrastructures and the Project of Europe*, 1-26. Basingstoke, Hampshire; New York: Palgrave Macmillan.
- Bailey, Richard W. (1969) Research Dictionaries, *American Speech*, 44(3): 166-72.
- Bailey, Richard W. (1973) Discussion Paper: Reflections on Technology in Lexicography, *Annals of the New York Academy of Sciences*,
- Bakker, Peter and Maarten Mous (eds.) (1994). *Mixed Languages: 15 Case Studies in Language Intertwining*. Amsterdam: Institute for Functional Research Into Language and Language Use (IFOTT).

- Barlow, Michael (2011) Corpus Linguistics and Theoretical Linguistics, *International Journal of Corpus Linguistics*, 16(1): 3-44.
- Barnhart, Clarence L (1973) Plan for a Central Archive for Lexicography in English, *Annals of the New York Academy of Sciences*, 211(1): 302-6.
- Basalla, George (1988). *The Evolution of Technology*. Cambridge [England]; New York: Cambridge University Press.
- Bataille, Georges (1929a) Informe, *Documents*, 1(7): 382.
- Bataille, Georges (1929b) Cheminée d'usine, *Documents*, 1(6): 329-32.
- Bates, Madeleine, et al. (1986). The IRUS transportable natural language database interface. *Proceedings from the first international workshop on Expert database systems*, 617-30. Redwood City, CA: Benjamin-Cummings Publishing.
- Bates, Marcia J (2002) Toward an Integrated Model of Information Seeking and Searching, *The New Review of Information Behaviour Research*, 3: 1-15.
- Béjoint, Henri (1994). *Tradition and Innovation in Modern English Dictionaries*. Oxford and New York: Clarendon Press; Oxford University Press.
- Béjoint, Henri (2000). *Modern Lexicography: An Introduction*. Oxford: Oxford University Press.
- Béjoint, Henri (2010). *The Lexicography of English*. Oxford: Oxford University Press.
- Ben-Hayyim, Zeiev (1966) A Hebrew Dictionary on Historical Principles, *ARIEL: A Review of the Arts and Sciences in Israel*, 13: 14-20.
- Bergenholtz, H. and S. Tarp (2002) Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen, *Lexicographica*, 18: 253-63.
- Bergenholtz, Hennig and Sandro Nielsen (2006) Subject-Field Components as Integrated Parts of LSP Dictionaries, *Terminology*, 12(2): 281-303.

- Bergenholtz, Henning and Sven Tarp (2003) Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions, *Hermes*, 31(2003): 171-96.
- Bergenholtz, Henning (2011). Access to and Presentation of Needs-Adapted Data in Monofunctional Internet Dictionaries. In Fuertes Olivera, Pedro and Henning Bergenholtz (eds.), *E-Lexicography: The Internet, Digital Initiatives and Lexicography*, 30-53. London and New York: Continuum.
- Bergenholtz, Henning and Inger Bergenholtz (2011). A Dictionary is a Tool, a Good Dictionary Is a Monofunctional Tool. In Fuertes Olivera, Pedro and Henning Bergenholtz (eds.), *E-Lexicography: The Internet, Digital Initiatives and Lexicography*, 187-207. London and New York: Continuum.
- Bernard, Pascale, et al. (2004) A computerized dictionary: Le Trésor de la langue Française informatisé (TLFi), *Proceedings of the Workshop on Enhancing and Using Electronic Dictionaries*, 40-3.
- Berners-Lee, T. (2006). *Linked Data-Design Issues*. 2009(09/20). <http://www.w3.org/DesignIssues/LinkedData.html>.
- Berry, David M. (2012). *Understanding Digital Humanities*. Houndmills, Basingstoke, Hampshire; New York: Palgrave Macmillan.
- Bierce, Ambrose (1906/2000). *The Unabridged Devil's Dictionary*. Athens & London: University of Georgia Press.
- Bittner, Johannes. (2003). *Digitalität, Sprache, Kommunikation: eine Untersuchung zur Medialität von digitalen Kommunikationsformen und Textsorten und deren varietätenlinguistischer Modellierung*. Berlin: E. Schmidt.
- Bockholt, Volker (1990). *Sprachmaterialkonzeptionen und ihre Realisierung in der kroatischen und serbischen Lexikographie*. Essen: Verl. Die Blaue Eule.
- Boguraev, Bran and Ted Briscoe (1989). Introduction. In Boguraev, Bran and Ted Briscoe (eds.), *Computational Lexicography for Natural Language Processing*, 1-40. London and New York: Longman.

- Boguraev, Branimir, et al. (1987) A Multi-Purpose Interface to an On-line Dictionary, *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics - EACL*, 63-9.
- Bolter, J. David (1991). *Writing Space: The Computer, Hypertext, and the History of Writing*. Hillsdale, N.J.: L. Erlbaum Associates.
- Bolter, J. David and Richard A. Grusin (1999). *Remediation: Understanding New Media*. Cambridge, Mass.: MIT Press.
- Bonin, Patrick (ed.) (2004). *Mental Lexicon: Some Words to Talk About Words*. Hauppauge, N.Y: Nova Science Publishers.
- Borin, Lars, et al. *Diabase: Towards a diachronic BLARK in support of historical studies*. <http://spraakbanken.gu.se/personal/lars/pblctns/lrec2010-diabase.pdf>. Accessed 15 March, 2010.
- Borin, Lars, et al. (2007) Naming the past: Named entity and animacy recognition in 19th century Swedish literature, *Workshop on Language Technology for Cultural Heritage Data*, 1-8.
- Boguraev, Bran and Ted Briscoe (1987) Large Lexicons for Natural Language Processing: Utilising the Grammar Coding System of LDOCE, *Computational Linguistics*, 13(3-4): 203-18.
- Bratley, Paul and Serge Lusignan (1976) Information Processing in Dictionary Making: Some Technical Guidelines, *Computers and the Humanities*, 10(3): 133-43.
- Brown, John Seely and Paul Duguid (1996). *The Social Life of Documents*. 1. <http://firstmonday.org/ojs/index.php/fm/article/view/466/387>. Accessed April 22, 2015.
- Brudney, James J. and Lawrence Baum (2013) Oasis or Mirage: The Supreme Court's Thirst for Dictionaries in the Rehnquist and Roberts Eras, *Fordham Law Legal Studies Research Paper*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2195644.

- Bugarski, Ranko. and Celia Hawkesworth (1992). *Language planning in Yugoslavia*. Columbus, Ohio: Slavica Publishers.
- Burkhanov, Igor (2003). Pragmatic specifications: Usage indications, labels, examples; dictionaries of style, dictionaries of collocations. In Sterkenburg, Piet Van Sterkenburg (ed.), *A Practical Guide to Lexicography*, 6, 102-11. Amsterdam; Philadelphia: John Benjamins Pub.
- Burkhanov, Igor (2004) Requirements for An" Ideal" Bilingual L1→ L2 Translation-Oriented Dictionary, *Lexikos*, 14: 17-34.
- Burkhart, Dagmar (2007) Ars memoriae und ars combinatoria in der Postmoderne. Am Beispiel von Milorad Pavic, *Die Welt der Slaven*, 40(2): 341-51.
- Busa, R. (1980) The annals of humanities computing: the Index Thomisticus, *Computers and the Humanities*, 14(2): 83-90.
- Butler, Thomas J. (1969) Jernej Kopitar's Role in the Serbian Language Controversy, *The Slavic and East European Journal*, 13(4): 488-479-88.
- Calzolari, Nicoletta (2008) Approaches Towards a Lexical Web: The Role of Interoperability, *Proceedings of the First International Conference on Global Interoperability for Language Resources*, 34-42.
- Cameron, Angus, et al. (eds.) (1970). *Computers and Old English Concordances*. Toronto: University of Toronto Press.
- Chapman, Robert L (1973) On Collecting for the Central Archive, *Annals of the New York Academy of Sciences*, 211(1): 307-11.
- Charlier, Robert (2005). Synergie und Konvergenz: Tradition und Zukunft historischer Semantik am Beispiel des Goethe- Wörterbuchs. In Müller, Ernst (ed.), *Begriffsgeschichte im Umbruch?*, 167-83.
- Chiarcos, Christian, et al. (eds.) (2012). *Linked Data in Linguistics. Representing and Connecting Language Data and Language Metadata*. Heidelberg: Springer.

- Chodorow, Martin S, et al. (1985). Extracting Semantic Hierarchies From a Large on-Line Dictionary. *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*, 299-304. Chicago: Association for Computational Linguistics.
- Christmann, R (2001) Books into bytes: Jacob and Wilhelm Grimm's Deutsches Wörterbuch on CD-ROM and on the Internet, *Literary and Linguistic Computing*, 16(2): 121-33.
- Christmann, R (2003) Towards the User: The Digital Edition of the Deutsche Wörterbuch by Jacob and Wilhelm Grimm, *Literary and linguistic computing*, 18(1): 11-22.
- Christopher, A. (Cal) Lee (2011) A framework for contextual information in digital collections, *Journal of Documentation*, 67(1): 95-143.
- Coblin, W. South (1993). Erh ya 爾雅. In Loewe, Michael. (ed.), *Early Chinese Texts: A Bibliographical Guide*, Berkeley, Calif.: Society for the Study of Early China: Institute of East Asian Studies, University of California, Berkeley.
- Considine, John (1998). Why Do Large Historical Dictionaries Give So Much Pleasure to Their Owners and Users? In Fontenelle, Thierry, et al. (eds.), *Actes EURALEX'98 Proceedings*, 579-87. Liège: University of Liège, English and Dutch Departments.
- Considine, John (2008). *Dictionaries in Early Modern Europe: Lexicography and the Making of Heritage*. Cambridge: Cambridge Univ Press.
- Cooper, Doug (2010) When Nice People Won't Share: Shy Data, Web APIs, and Beyond, *Second International Conference on Global Interoperability for Language Resources (ICGL 2010)*, n. pag.
- Cotkin, George (1996) "Hyping the Text": Hypertext, Postmodernism, and the Historian, *American Studies*, 37(2): 103-16.
- de Schryver, Gilles-Maurice (2003) Lexicographer's Dreams in the Electronic-Dictionary Age, *International Journal of Lexicography*, 16 (2): 143-99.

- de Tollenaere, F (1971) La table ronde des dictionnaires historiques (Florence 3-5 mai 1971), *Cahiers de Lexicologie*, 19: 116-28.
- De Tollenaere, F (1972) Encoding Techniques in Dutch Historical Lexicography, *Computers and the Humanities*, 6(3): 147-52.
- Delany, Paul. and George P. Landow (1991). Hypertext, Hypermedia and Literary Studies. In Delany, Paul. and George P. Landow (eds.), *Hypermedia and Literary Studies*, 3-52. Cambridge, Mass.: MIT Press.
- Deleuze, Gilles and Félix Guattari (1987). *A Thousand Plateaus: Capitalism and Schizophrenia*. Minneapolis: University of Minnesota Press.
- Deleuze, Gilles and Claire. Parnet (1987). *Dialogues*. New York: Columbia University Press.
- Deleuze, Gilles and Félix Guattari (1994). *What Is Philosophy?* New York: Columbia University Press.
- Delić, Jovan (1990). Die Rezeption des Kosovo-Mythos in Vuks *Serbischem Wörterbuch*. In Potthoff, Wilfried (ed.), *Vuk Karadžić im europäischen Kontext*, 31-78. Heidelberg: Carl Winter Universitätsverlag.
- Derrida, Jacques. (1997). *Of Grammatology*. Baltimore: Johns Hopkins University Press.
- Desproges, Pierre (1985). *Dictionnaire superflu à l'usage de l'élite et des bien nantis*. Paris: Editions du Seuil.
- Dickey, Eleanor (2007). *Ancient Greek Scholarship: A Guide to Finding, Reading, and Understanding Scholia, Commentaries, Lexica, and Grammatical Treatises: From Their Beginnings to the Byzantine Period*. Oxford and New York: Oxford University Press.
- Dickey, Eleanor (2010). Greek dictionaries ancient and modern. In Stray, Christopher (ed.), *Classical Dictionaries: Past, Present and Future*, 5-24. London: Duckworth.
- Dickinson, William (1879). *A Glossary of Words and Phrases Pertaining to the Dialect of Cumberland*. London: English Dialect Society.

- Dillon, Andrew (2001) Beyond Usability: Process, Outcome and Affect in Human-Computer Interactions, *Canadian Journal of Library and Information Science*, 26(4): 57-69.
- Dolezal, Fredric and Don R McCreary (1999). *Pedagogical Lexicography Today: A Critical Bibliography on Learners' Dictionaries with Special Emphasis On Language Learners and Dictionary Users*. Tübingen: Niemeyer.
- Dolezal, Fredric (2009). World Englishes and Lexicography. In Kachru, Braj B., et al. (eds.), *The Handbook of World Englishes*, Blackwell handbooks in linguistics, 694-708. Malden, MA; Oxford: Blackwell.
- Driscoll, M.J. (2006). Levels of Transcription. In Burnard, Lou, et al. (eds.), *Electronic Textual Editing*, 254-61. New York: Modern Language Association.
- Dronjic, Vedran (2011) Serbo-Croatian: The making and breaking of an Ausbausprache, *Language Problems & Language Planning*, 35(1): 1-14.
- Dubois, Jean (1970) Dictionnaire et discours didactique, *Langages*, 5(19): 35-47.
- Duro, Aldo (1966) Les nouvelles méthodes du dictionnaire historique de la langue italienne, *Cahiers de Lexicologie*, 8: 95-111.
- Duro, Aldo (1968) Humanities Computing Activities in Italy, *Computers and the Humanities*, 3: 49-52.
- Duro, Aldo (1973). Élaborations électroniques de textes effectuées par l'Accademia della Crusca, pour la préparation du Dictionnaire historique de la langue italienne. In Zampolli, Antonio (ed.), *Linguistica matematica e calcolatori. Atti del Convegno e della prima Scuola internazionale*. Pisa, 16 VIII/6 IX 1970, 53-75. Firenze: Leo S. Olschki.
- Earhart, Amy E (2012) The Digital Edition and the Digital Humanities, *Textual Cultures: Texts, Contexts, Interpretation*, 7(1): 18-28.
- Edmonds, Philip and Graeme Hirst (2002) Near-Synonymy and Lexical Choice, *Computational Linguistics*, 28(2): 105-44.

- Edward Foster, Allen and David Ellis (2014) Serendipity and Its Study, *Journal of Documentation*, 70(6): 1015-38.
- Eisenstein, Elizabeth L. (1979). *The Printing Press as an Agent of Change: Communications and Cultural Transformations in Early Modern Europe*. Cambridge; New York: Cambridge University Press.
- Ellingson, Terry Jay. (2001). *The Myth of the Noble Savage*. Berkeley: University of California Press.
- Epstein, Mikhail (2012). *The Transformative Humanities: A Manifesto*. New York: Bloomsbury.
- Erdelez, Sanda (1997) Information Encountering: A Conceptual Framework for Accidental Information Discovery, Information Seeking in Context: Proceedings of an International Conference on Research in Information Needs, Seeking and Use in Different Contexts, 14-16 August 1996, Tampere, Finland: 412-21.
- Erdelez, Sanda (2005). Information Encountering. In Fisher, Karen E., et al. (eds.), *Theories of Information Behavior*, 179-84. Medford, N.J.: Published for the American Society for Information Science and Technology by Information Today.
- Erl, Thomas, et al. (2017). *Service-Oriented Architecture: Analysis and Design for Services and Microservices*. Boston: Prentice Hall.
- Eschker, Wolfgang (ed.) (1988). *Jacob Grimm und Vuk Karadžić: Zeugnisse einer Gelehrtenfreundschaft*. Kassel: E. Röth-Verlag.
- Evens, Martha W. (ed.) (1988). *Relational Models of the Lexicon: Representing Knowledge in Semantic Networks*. Cambridge: Cambridge University Press.
- Faraj, Samer and Bijan Azad (2012) The materiality of technology: An affordance perspective, *Materiality and organizing: Social interaction in a technological world*, 237-58.

- Febvre, Lucien Paul Victor and Henri-Jean Martin (1958/2010). *The Coming of the Book: The Impact of Printing, 1450-1800*. London: Verso.
- Fellbaum, Christiane (ed.) (2009). *Idioms and Collocations: Corpus-based Linguistic and Lexicographic Studies*. London: Continuum.
- Ferguson, Charles (1959) Diglossia, *Word*, 15: 325-40.
- Fillmore, Charles J. (1988) Keynote Lecture, *British Association for Applied Linguistics*,
- Finkenstaedt, Thomas, et al. (1970). *A Chronological English Dictionary. Listing 80000 Words in Order of Their Earliest Known Occurrence*. Heidelberg: C. Winter.
- Finkelkraut, Alain (1981). *Petit dictionnaire illustré: Les mots qui manquent au dico*. Paris: Seuil.
- Fishman, Joshua A. (1995). Dictionary as Culturally Constructed and as Culture-Constructing Artifacts: The Reciprocity View as Seen from Yiddish Sources. In Kachru, Braj B. and Ladislav Zgusta (eds.), *Cultures, Ideologies, and the Dictionary: Studies in Honor of Ladislav Zgusta*, 29-34. Tübingen: Niemeyer.
- Flaubert, Gustave (1913/2002). *Dictionnaire des idées reçues*. Paris: Éditions du Boucher.
- Fomin, Maxim and Gregory Toner (2006) Digitizing a Dictionary of Medieval Irish: the eDIL Project, *Literary and Linguistic Computing*, 21(1): 83.
- Fournier, Johannes (2001) New directions in Middle High German lexicography: dictionaries interlinked electronically, *Literary and Linguistic Computing*, 16(1): 99-111.
- Frawley, William (1985) Intertextuality and the dictionary: Toward a deconstructionist account of lexicography, *Dictionaries: Journal of the Dictionary Society of North America*, 7: 1-20.
- Friedman, Victor A (1997) One grammar, three lexicons: ideological overtones and underpinnings in the Balkan Sprachbund, *Chicago Linguistic Society*, 33: 23-44.

- Fuertes Olivera, Pedro and Henning Bergenholtz (2011). Introduction: The Construction of Internet Dictionaries. In Fuertes Olivera, Pedro and Henning Bergenholtz (eds.), *E-Lexicography: The Internet, Digital Initiatives and Lexicography*, 1-16. London and New York: Continuum.
- Gabler, Hans Walter (2006). Moving a Print-Based Editorial Project Into Electronic Form. In Burnard, Lou, et al. (eds.), **Electronic Textual Editing**, 339-45. New York: Modern Language Association of America.
- Gailey, Amanda and Andrew Jewell (2012) Editors' Introduction to the First Issue of Scholarly Editing: The Annual of the Association for Documentary Editing, *Scholarly Editing*, 33: 1-7.
- Galison, Peter (1997). *Image and Logic: A Material Culture of Microphysics*. Chicago: University of Chicago Press.
- Geeraerts, Dirk (1996). Entfernung, Fortschritt und Unvollständigkeit: das historische Verhältnis von Lexikologie und Lexikographie. *Linguistische Theorie und lexicographische Praxis*, 7-15. Tübingen: Max Niemeyer.
- Geertz, Clifford. (1973). Thick Description: Toward an Interpretive Theory of Culture. *The Interpretation of Cultures: Selected Essays*, 3-30. New York: Basic Books.
- Genette, Gérard (1997). *Paratexts: Thresholds of Interpretation*. Cambridge and New York,: Cambridge University Press.
- Gibson, James J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Glinister, Fay, et al. (2007) Introduction. Verrius, Festus and Paul, *Bulletin of the Institute of Classical Studies*, 50(S93): 1-9.

- Gloning, Thomas and R. Welter (2001). Wortschatzarchitektur und elektronische Wörterbücher. Goethes Wortschatz und das Goethe-Wörterbuch. In Lemberg, Ingrid, et al. (eds.), *Chancen und Perspektiven computergestützter Lexikographie: Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher*, 107, 117-32. Tübingen: M. Niemeyer.
- Gold, Matthew K. (ed.) (2012). *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press.
- Goldfarb, Charles F, et al. (1970) An Online System for Integrated Text Processing, *Proceedings of the American Society for Information Science*, 7: 147-50.
- Goldfarb, Charles F (1981) A Generalized Approach to Document Markup, *ACM Sigplan Notices*, 16(6): 68-73.
- Goldfarb, Charles F and Yuri Rubinsky (1990). *The SGML Handbook*. Oxford: Oxford University Press.
- Gorman, M.E. (2002) Levels of Expertise and Trading Zones: A Framework for Multidisciplinary Collaboration, *Social Studies of Science*, 32(5/6): 933-8.
- Gorman, Michael E. (2010). *Trading Zones and Interactional Expertise: Creating New Kinds of Collaboration*. Cambridge, Mass.: MIT Press.
- Gouws, Rufus H. and D J Prinsloo (1998) Cross-Referencing as a Lexicographic Device, *Lexikos*, 8(1): 17-36.
- Gouws, Rufus H. (2011). Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries. In Fuertes Olivera, Pedro and Henning Bergenholtz (eds.), *E-Lexicography: The Internet, Digital Initiatives and Lexicography*, 17-29. London and New York: Continuum.
- Gray, Jack C. (1986) Creating the Electronic New Oxford English Dictionary, *Computers and the Humanities*, 20: 45-9.
- Grčević, Mario (2009) Jernej Kopitar kao strateg Karadžićeve književnojezične norme, *Filologija*, 53: 1-54.

- Greenberg, Marc L. (2011). The Illyrian Movement: A Croatian Vision of South Slavic Unity. In Fishman, Joshua A and Ofelia García (eds.), 364-80. Oxford and New York: Oxford University Press, USA.
- Greenberg, Robert D (1996) The Politics of Dialects Among Serbs, Croats, and Muslims in the Former Yugoslavia, *East European Politics & Societies*, 10(3): 393-415.
- Grimm, Jacob (1815/1869). Мала простионародна славно-сербска Песнарица, издана Вуком Стефановићем (d. i. kleines serben-slawisches volksierbuch, herausgegeben durch Wuk Stephanowitsch (Wolf Stephanson). *Recensionen und vermischte Aufsätze. Erster Theil*, 427-36. Berlin: Ferd. Dümmlers Verlagsbuchhandlung.
- Hanks, Patrick (2008). Do Word Meanings Exist? In Fontenelle, Thierry (ed.), *Practical Lexicography: A Reader*, 125-34. Oxford, New York: Oxford University Press.
- Harris, Roy and Christopher Hutton (2007). *Definition in Theory and Practice: Language, Lexicography and the Law*. London; New York: Continuum.
- Hartmann, Frank (2003). *Mediologie: Ansätze einer Medientheorie der Kulturwissenschaften*. Wien: WUV.
- Hartmann, R R K and Gregory James (1998). *Dictionary of Lexicography*. London; New York: Routledge.
- Harwood, Nigel (2002) Taking a lexical approach to teaching: principles and problems, *International Journal of Applied Linguistics*, 12(2): 139-55.
- Haslam, Michael W (1994a) The Homer Lexicon of Apollonius Sophista: I. Composition and Constituents, *Classical philology*, 89(1): 1-45.
- Haslam, Michael W (1994b) The Homer Lexicon of Apollonius Sophista: II. Identity and Transmission, *Classical Philology*, 89(2): 107-19.
- Hass Weinberg, Bella (2010). Indexing: History and Theory. In Bates, Marcia J. and Mary Niles Maack (eds.), *Encyclopedia of Library and Information Sciences*, Boca Raton, FL: CRC Press.

- Haß, Ulrike (2011) Chancen und Perspektiven der historischen Lexikografie des Deutschen, *Lexicographica*, 27: 45-61.
- Haugen, Einar (1966) Dialect, Language, Nation, *American Anthropologist New Series*, 68(4): 922-35.
- Hausmann, Franz Josef (1977). *Einführung in die Benutzung der neufranzösischen Wörterbücher*. Tübingen: Niemeyer.
- Hausmann, Franz Josef (1989a). Die gesellschaftlichen Aufgaben der Lexikographie in Geschichte und Gegenwart. In Hausmann, Franz Josef, et al. (eds.), *Wörterbücher : ein internationales Handbuch zur Lexikographie*, 1-18. Berlin; New York: W. de Gruyter.
- Hausmann, Franz Josef (1989b). Kleine Weltgeschichte der Metalexikographie. *Wörterbücher in der Diskussion. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*, 75-109. Tübingen: Niemeyer.
- Hausmann, Franz Josef and Herbert Ernst Wiegand (1989). Component Parts and Structures of General Monolingual Dictionaries: A Survey. In Hausmann, Franz Josef, et al. (eds.), *Wörterbücher : ein internationales Handbuch zur Lexikographie*, Berlin; New York: W. de Gruyter.
- Hausmann, Franz Josef (1989c). Die Markierung im allgemeinen einsprachigen Wörterbuch: eine Übersicht. In Hausmann, Franz Josef, et al. (eds.), *Wörterbücher: ein internationales Handbuch zur Lexikographie*, 1, 649-57. Berlin; New York: W. de Gruyter.
- Hayashi, Tetsuro (1978). *The Theory of English Lexicography, 1530-1791*. Amsterdam: John Benjamins.
- Hayashi, Yoshihiko and Toru Ishida (2006) A dictionary model for unifying machine readable dictionaries and computational concept lexicons, *LREC 2006*,
- Hayles, N Katherine (1997) Corporeal Anxiety in Dictionary of the Khazars: What Books Talk About in the Late Age of Print When They Talk About Losing Their Bodies, *MFS Modern Fiction Studies*, 43(3): 800-20.

- Hayles, N Katherine (2003) Translating Media: Why We Should Rethink Textuality, *The Yale Journal of Criticism*, 16(2): 263-90.
- Heid, Ulrich (2011). Electronic Dictionaries as Tools: Toward an Assessment of Usability. In Fuertes Olivera, Pedro and Henning Bergenholtz (eds.), *E-Lexicography: The Internet, Digital Initiatives and Lexicography*, 287-304. London and New York: Continuum.
- Heide, Lars (2009). *Punched-Card Systems and the Early Information Explosion, 1880-1945*. Baltimore: The John Hopkins University Press.
- Heidegger, Martin (1962). *Being and Time*. London: SCM Press.
- Heidorn, George E., et al. (1982) The EPISTLE Text-Critiquing System, *IBM Systems Journal*, 21(3): 305-26.
- Henderson, Michael M. T. (1974) Processing Data for DARE: Current Practices, *American Speech*, 49(1/2): 119-22.
- Herder, Johann Gottfried (1772). *Abhandlung über den Ursprung der Sprache, welche den von der Königl. Academie der Wissenschaften für das Jahr 1770 gesetzten Preis erhalten hat: Auf Befehl der Academie herausgegeben*. Berlin: Christian Friedrich Voss.
- Hockey, Susan M (2000). Dictionaries and Lexcial Databases. *Electronic texts in the humanities: Principles and Practice*, 146-71. Oxford & New York: Oxford University Press.
- Homer, Frederic D and Charles H Levine (1985) Triviocracy: Sayre's Law Revisited, *Review of Policy Research*, 5(2): 241-52.
- Hoover, David (2005) Hot-Air Textuality: Literature after Jerome McGann, *Text Technology*, 14(2): 71-103.
- Horsley, Ross (2011). *My First Dictionary: Corrupting Young Minds One Word at a Time*. New York: ItBooks.
- Hüllen, Werner and Rainer Schulze (1988). *Understanding the Lexicon: Meaning, Sense, and World Knowledge In Lexical Semantics*. Tübingen: M. Niemeyer.

- Hüllen, Werner (1999). *English Dictionaries, 800-1700: The Topical Tradition*. Oxford [England] New York: Clarendon Press Oxford University Press.
- Huntsman, Jeffrey F (1978) Computers and Medieval English Lexicography, *Computers and the Humanities*, 12(1-2): 53-60.
- Hutchby, I. (2001) Technologies, Texts and Affordances, *Sociology*, 35(2): 441-56.
- Ide, Nancy, et al. (2007). A Formal Model of Dictionary Structure and Content. *Information Technology Research Institute Technical Report Series*, Brighton: University of Brighton.
- Ide, Nancy M and Jean Véronis (1995). *Text Encoding Initiative: Background and Contexts*. Cambridge, MA: The MIT Press.
- IFLA (1998). *Functional Requirements for Bibliographic Records: Final Report*. <https://archive.ifla.org/VII/s13/frbr/frbr1.htm#3.2>. Accessed September 23, 2016.
- Ilić Marković, Gordana, et al. (eds.) (2012). *An den Anfängen der serbischen Philologie. Salo debeloga jera libo azbukoprotres von Sava Mrkalj (1810-2010)*. Frankfurt am Main: Peter Lang.
- Ingria, Robert J. P. (1995). Lexical Information for Parsing Systems: Points of Convergence and Divergence. In Walker, Donald E., et al. (eds.), *Automating the Lexicon: Research and Practice in a Multilingual Environment*, 93-170. Oxford: Oxford University Press.
- Innis, Harold Adams (1951/1995). *The Bias of Communication*. Toronto: University of Toronto Press.
- Iveković, F. and Ivan Broz (1901). *Rječnik hrvatskoga jezika*. Zagreb: Štamparija Kalra Albrehta (Jos. Wittasek).
- JAZU (1880). *Rječnik hrvatskoga ili srpskoga jezika*. Zagreb: Jugoslavenska akademija znanosti i umjetnosti.
- Johansson, Stig (1985) Word Frequency and Text Type: Some Observations Based on the LOB Corpus of British English Texts, *Computers and the Humanities*, 19(1): 23-36.

- Johnson, Samuel (1785) *A Dictionary of the English Language: In Which the Words Are Deduced From Their Originals, and Illustrated in Their Different Significations by Examples From the Best Writers: To Which Are Prefixed, a History of the Language, and an English Grammar*. Edited by London: J. F. and C Rivington et al.
- Johnson, Samuel (2017). *The Plan of a Dictionary of the English Language (1747)*. In Shapiro, Rebecca (ed.), *Fixing Babel: An Historical Anthology of Applied Lexicography, 176-200*. Lewisburg and Lanham, Maryland: Bucknell University Press and The Rowman & Littlefield Publishing Group, Inc.
- Joseph, John Earl. (1987). *Eloquence and Power: The Rise of Language Standards and Standard Languages*. London: Frances Pinter.
- Kahane, Henry and Renée Kahane (1992). *The Dictionary as Ideology. Sixteen Case Studies*. In Zgusta, Ladislav (ed.), *History, Languages, and Lexicographers*, 19-76. Tübingen: Niemeyer.
- Kamp, Hermann (1969) *Methoden zur Herstellung und Auswertung von Wörterbüchern mit Hilfe der elektronischen Datenverarbeitung, Niederdeutsches Wort*, 9: 73-96.
- Kay, Michael (2008). *XSLT 2.0 and XPath 2.0: Programmer's Reference*. Indianapolis, IN: Wiley Pub.
- Kazman, Rick (1986). *Structuring the Text of the Oxford English Dictionary Through Finite State Transduction*. Waterloo, Ontario: University of Waterloo, Computer Science Department.
- Keating, John G., et al. (2010). *A Digital Edition of a Spanish 18th Century Account Book: Part 1 - User Driven Digitisation*. In Ryder, Sean and Malte Rehbein (eds.), *Jahrbuch für Computerphilologie*, 169-88. Paderborn: Mentis Verlag.
- Kelly-Bootle, Stan (1995). *The Computer Contradictionary*. Cambridge, Mass.: MIT Press.
- Kenyon, John Samuel and Thomas A. Knott (1953). *A Pronouncing Dictionary of American English*. Springfield, Mass.: G. & C. Merriam Co.

- Keseling, Gisbert, et al. (1970) Richtlinien zur Ablochung und zentralen Speicherung mundartlichen Wortmaterials des Deutschen, *Germanistische Linguistik*, 2(70): 179-242.
- Kilgarriff, Adam (2008). "I Don't Believe in Word Senses". In Fontenelle, Thierry (ed.), *Practical Lexicography: A Reader*, 135-51. Oxford, New York: Oxford University Press.
- Kinable, Dirk (2006) Computerized Restoration of Historical Dictionaries: Uniformization and Date-assigning in Dictionary Quotations of the Woordenboek der Nederlandsche Taal, *Literary and Linguistic Computing*, 21(3): 295-310.
- Kirkness, Alan (2008) Digitalisierung - Vernetzung - Europäisierung: Zur Zukunft der historischen Lexikographie des Deutschen, *Lexicographica*, 2007: 7-38.
- Kirkpatrick, Betty (1985). A Lexicographical Dilemma: Monolingual Dictionaries for the Native Speaker and for the Learner. In Ilson, Robert (ed.), *Dictionaries, Lexicography, and Language Learning*, 7-13. Oxford; New York: Pergamon Press in association with the British Council.
- Klein, Devorah E and Gregory L Murphy (2001) The Representation of Polysemous Words, *Journal of Memory and Language*, 45(2): 259-82.
- Klepousniotou, Ekaterini, et al. (2008) Making Sense of Word Senses: The Comprehension of Polysemy Depends on Sense Overlap, *Journal of experimental psychology. Learning, memory, and cognition*, 34(6): 1534.
- Knowles, Francis E (1989). Computers and Dictionaries. In Hausmann, Franz Josef, et al. (eds.), *Wörterbücher: ein internationales Handbuch zur Lexikographie*, 1, 1645-72. Berlin; New York: W. de Gruyter.
- Koffka, Kurt (1935). *Principles of Gestalt Psychology*. New York: Harcourt, Brace and Company.
- Kosem, Iztok, et al. (2018) The Image of the Monolingual Dictionary Across Europe. Results of the European Survey of Dictionary Use and Culture, *International Journal of Lexicography*, 32(1): 92-114.

- Kramer, Johannes (1996). I glossari tardo-antichi di tradizione papiracea. In Hamesse, J (ed.), *Les manuscrits des lexiques et glossaires de l'Antiquité tardive à la fin du moyen âge: Actes du Colloque international (Erice, 23-30 septembre 1994)*, 23-55.
- Kramer, Johannes (2011). *Vulgärlateinische Alltagsdokumente auf Papyri, Ostraka, Täfelchen und Inschriften*. Walter de Gruyter.
- Krauss, Rosalind (1985) *Corpus delicti*, *October*, 33: 31-72.
- Kretschmer, Anna (1994) Und noch einmal zur Diglossie, *Wiener Slawistischer Almanach*, 33: 181-94.
- Kretschmer, Anna (2004) Vom Kirchenslavischen zu Schriftsprachen der Neuzeit (Schriftserbisch um 1700), *Welt der Slaven*, 49: 145-60.
- Kretschmer, Anna (2008). Slavenoserbisch – Zeitalter der Aufklärung im serbischen Kulturparadigma. In Dahmen, Wolfgang and Petra Ressel Himstedt-Vaid, Gerhard (eds.), *Grenzüberschreitungen: Traditionen und Identitäten in Südosteuropa; Festschrift für Gabriella Schubert*, 314-24. Wiesbaden: Harrassowitz Verlag.
- Kromann, Hans-Peder, et al. (1984). Überlegungen zu Grundfragen der zweisprachigen Lexikographie. In Wiegand, Herbert Ernst (ed.), *Studien zur neuhochdeutschen Lexikographie*, 159-238. Hildesheim and New York: Olms.
- Kučera, Henry and Winthrop Nelson Francis (1967). *Computational Analysis of Present-Day American English*. Providence, RI: Dartmouth Publishing Group.
- Kunzmann-Müller, Barbara (ed.) (2000). *Die Sprachen Südosteuropas heute: Umbrüche und Aufbruch*. Berliner slawistische Arbeiten, Bd. 12. Frankfurt am Main; New York: Lang.
- Kuutti, Kari (1995). Activity theory as a potential frame- work for human-computer interaction research. In Nardi, Bonnie A. (ed.), *Context and Consciousness: Activity Theory and Human-Computer Interaction*, 17-44. Cambridge, MA: The MIT Press.

- Landau, Sidney I. (1984). *Dictionaries: The Art and Craft of Lexicography*. New York: The Scribner Press.
- Landow, George P. (1994). *Hyper/text/theory*. Baltimore; London: Johns Hopkins University Press.
- Lawley, James and Penny Tompkins (2008). *Maximising Serendipity: The Art of Recognising and Fostering Potential: A Systemic Approach to Change*. <http://www.cleanlanguage.co.uk/articles/articles/224/1/Maximising-Serendipity/Page1.html>. Accessed June 25, 2017.
- Lefler, Anna (2011). *The Chicktionary: From a-Line to Z-Snap, the Words Every Woman Should Know*. Avon, Mass.: Adams Media.
- Lehiste, Ilse and Pavle Ivić (1986). *Word and Sentence Prosody in Serbocroatian*. Cambridge, Mass.: MIT Press.
- Lehmann, W P (1973) On the Design of a Central Archive for Lexicography in English, *Annals of the New York Academy of Sciences*, 211(1): 312-7.
- Leiris, Michel (1925a) Glossaire: j'y serre mes gloses, *La révolution surréaliste*, 3: 6-7.
- Leiris, Michel (1925b) Glossaire: j'y serre mes gloses, *La révolution surréaliste*, 4: 20-1.
- Leiris, Michel (1925c) Glossaire: j'y serre mes gloses, *La révolution surréaliste*, 6: 20-1.
- Leiris, Michel (1939). *Glossaire j'y serre mes gloses*. Paris: Éditions de la Galerie Simon.
- Leiris, Michel (1969). *Mots sans mémoire*. Paris: Gallimard.
- Leitner, A (1998) Dictionary of the Khazars as an Epistemological Metaphor, *Review of Contemporary Fiction*, 18: 155-63.
- Lemberg, I, et al. (1998) Der Weg des Deutschen Rechtswörterbuchs in das Internet, *Wörterbücher in der Diskussion III. Vorträge aus dem Heidelberger Lexikographischen Kolloquium, Tübingen: Niemeyer*, 262-84.
- Lemnitzer, Lothar, et al. (2010) Representing Human and Machine Dictionaries in Markup languages (SGML, XML), <http://arxiv.org/pdf/0912.2881>.

- Leonardi, Paul M and Stephen R Barley (2010) What's Under Construction Here? Social Action, Materiality, and Power in Constructivist Studies of Technology and Organizing, *Academy of Management Annals*, 4(1): 1-51.
- Lewin, Kurt (1969). *Grundzüge der topologischen Psychologie*. Bern, Stuttgart: Hans Huber.
- Leyerle, John (1971) "The Dictionary of Old English": A Progress Report, *Computers and the Humanities*, 5(5): 279-83.
- Liptak, Adam (2011). *Justices Turning More Frequently to Dictionary, and Not Just for Big Words*. http://www.nytimes.com/2011/06/14/us/14bar.html?_r=0. Accessed June 13, 2011.
- Lisac, Josip (2003). *Hrvatski dijalekti i govori štokavskog narječja i hrvatski govori torlačkog narječja*. Zagreb: Tehnička knjiga.
- Lobenstein-Reichmann, Anja (1998). *Freiheit bei Martin Luther: lexikographische Textanalyse als Methode historischer Semantik*. Berlin [u.a.]: de Gruyter.
- Lobenstein-Reichmann, Anja (2000). Dichtung und Wahrheit in der Geschichtsschreibung. Lexikographie im Dienste historischer Erkenntnisfindung, dargestellt am Beispiel Luthers. In Wiegand, Herbert Ernst (ed.), *Wörterbücher in der Diskussion IV. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*, 175-202. Tübingen: Niemeyer.
- Lobenstein-Reichmann, Anja (2007) Allgemeine Überlegungen zur Retrodigitalisierung historischer Wörterbücher des Deutschen, *Lexicographica*, 2007: 173-98.
- Longinovic, Tomislav (1998) Chaos, Knowledge, and Desire: Narrative Strategies in Dictionary of the Khazars, *Review of Contemporary Fiction*, 18(2): 183-90.
- Lord, Albert Bates. (2000). *The Singer of Tales*. Cambridge, Mass.: Harvard University Press.
- Louw, Bill (1993) Irony in the Text or Insincerity in the Writer, *Text and Technology: In Honour of John Sinclair*, 157-76.

- Lukežić, Iva (1990). *Čakavski ikavsko-ekavski dijalekt*. Rijeka: Izdavački Centar.
- Lynch, Clifford (2002) Digital collections, digital libraries and the digitization of cultural heritage information, *First Monday*, 7(5-6):
- Lynch, Jack (2009). *The Lexicographer's Dilemma: The Evolution of "Proper" English, from Shakespeare to South Park*. New York: Walker & Co.
- Lynch, Jack (2016). *You Could Look It Up : The Reference Shelf From Ancient Babylon To Wikipedia*. New York: Bloomsbury Press.
- Mackintosh, Kristen (2006). Biased Books by Harmless Drudges: How Dictionaries Are Influenced by Social Values. In Bowker, Lynne (ed.), *Lexicography, Terminology, and Translation: Text-Based Studies in Honour of Ingrid Meyer*, 45-63. Ottawa: University of Ottawa Press.
- Magner, Thomas and Milena Marić (2002) Bosnian: The Crafting of a Language, *Geolinguistics*, 28: 55-65.
- Makkai, A. (1980). Theoretical and Practical Aspects of an Associative Lexicon for 20th-Century English. In Zgusta, Ladislav (ed.), *Theory and Method In Lexicography: Western and Non-Western Perspectives*, Columbia, S. Carolina: Hornbeam Press.
- Makri, Stephann and Ann Blandford (2012a) Coming Across Information Serendipitously—Part 1: A Process Model, *Journal of Documentation*, 68(5): 684-705.
- Makri, Stephann and Ann Blandford (2012b) Coming Across Information Serendipitously—Part 2: A Classification Framework, *Journal of Documentation*, 68(5): 706-24.
- Markus, M Lynne and Mark S Silver (2008) A Foundation for the Study of IT Effects: A New Look at DeSanctis and Poole's Concepts of Structural Features and Spirit, *Journal of the Association for Information Systems*, 9 (10/11): 609-32.

- Marmaridou, Sophia (2000). *Pragmatic Meaning and Cognition*. Amsterdam; Philadelphia: J. Benjamins Pub. Co.
- Matras, Yaron and Peter Bakker (2003). *The mixed language debate : theoretical and empirical advances*. Berlin; New York: Mouton de Gruyter.
- McArthur, T. B. (1986). Thematic Lexicography. In Hartmann, R. R. K (ed.), *The History of Lexicography: Papers from the Dictionary Research Centre Seminar at Exeter, March 1986*, 40, Amsterdam; Philadelphia: J. Benjamins.
- McCarty, Willard (1996) Finding Implicit Patterns in Ovid's "Metamorphoses" with "Tact", *CH Working Papers*, 1(1):
- McCarty, Willard (2002). Network with a thousand entrances: commentary in an electronic age? In Gibson, Roy K. and Christina Shuttleworth Kraus (eds.), *The Classical Commentary: History, Practices, Theory*, 359-402. Leiden: Brill.
- McCarty, Willard (2007). Knowing.: Modeling in Literary Studies. In Siemens, Raymond George and Susan Schreibman (eds.), *A Companion to Digital Literary Studies*, 391-401. Malden, MA: Blackwell Publishing.
- McCay-Peet, Lori and Elaine G Toms (2010) The Process of Serendipity in Knowledge Work, *Proceedings of the Third Symposium on Information Interaction in Context*: 377-82.
- McDavid Jr., Raven I. (1973) Opening Remarks, *Annals of the New York Academy of Sciences*, 211(1): 2-7.
- McGann, Jerome J. (2001). *Radiant Textuality: Literature After the World Wide Web*. New York: Palgrave.
- McLennan, Sean (1996) Sociolinguistic Analysis of 'Serbo-Croatian', *Calgary Working Papers in Linguistics*, 18: 103-9.
- McLuhan, Marshall (1962). *The Gutenberg Galaxy: The Making of Typographic Man*. Toronto: University of Toronto Press.

- Meijs, Willem (1992). Computers and Dictionaries. In Butler, Christopher (ed.), *Computers and Written Texts*, Applied Language Studies, 141-65. Oxford, UK & Cambridge, USA: Blackwell.
- Michaels, Claire F and Claudia Carello (1981). *Direct Perception*. Prentice-Hall Englewood Cliffs, NJ.
- Michiels, Archibal, et al. (1980) Exploiting a Large Data Base by Longman, *Proceedings of the 8th Conference on Computational Linguistics*, 374-82.
- Miller, George A. (1985). Wordnet: A Dictionary Browser. *Information in Data. Proceedings of the First Conference of the UW Centre for the New Oxford Dictionary*, 25-8. Waterloo, Canada: University of Waterloo.
- (1993). *Five Papers on WordNet*. <ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps>. Accessed 1 December 2008,
- MLA (2011). *Guidelines for Editors of Scholarly Editions*. Modern Language Association. <https://www.mla.org/Resources/Research/Surveys-Reports-and-Other-Documents/Publishing-and-Scholarship/Reports-from-the-MLA-Committee-on-Scholarly-Editions/Guidelines-for-Editors-of-Scholarly-Editions>. Accessed January, 28, 2017.
- MLA (2015). *Considering the Scholarly Edition in the Digital Age: A White Paper of the Modern Language Association's Committee on Scholarly Editions*. <https://scholarlyeditions.mla.hcommons.org/2015/09/02/cse-white-paper/>. Accessed May 23, 2017.
- Mooijaart, Marijke and Marijke van der Wal (eds.) (2009). *Yesterday's Words: Contemporary, Current and Future Lexicography*. Newcastle: Cambridge Scholars Publishing.
- Morrissey, Robert (1993) Texts and Contexts: The ARTFL Database in French Studies, *Profession*, 27-33.
- Moulin, C, et al. (eds.) (2011). *Research Infrastructures in the Digital Humanities*. Strasbourg: European Science Foundation.

- Müller-Spitzer, Carolin (2005). Die Modellierung lexikografischer Daten und ihre Rolle im lexikographischen Prozess. In Hass, Ulrike (ed.), *Grundfragen der elektronischen Lexikographie : Elexiko, das Online-Informationssystem zum deutschen Wortschatz*, 12, 21-54. Berlin; New York: W. de Gruyter.
- Müller-Spitzer, Carolin (2011) Der Aufbau einer maßgeschneiderten XML-basierten Modellierung für ein Wörterbuchnetz, *Datenmodellierung für Internetwörterbücher Datenmodellierung für Internetwörterbücher*, 37.
- Müller-Spitzer, Carolin (ed.) (2014). *Using Online Dictionaries*. Berlin: De Gruyter.
- Müller-Spitzer, Carolin, et al. (2015) Observing Online Dictionary Users: Studies Using Wiktionary Log Files, *International Journal of Lexicography*, 28(1): 1-26.
- Murphy, M. Lynne (2003). *Semantic Relations and the Lexicon : Antonymy, Synonymy, and Other Paradigms*. Cambridge, UK; New York: Cambridge University Press.
- Murray, Katherine Maude Elisabeth (2001). *Caught in the Web of Words: James A.H. Murray and the Oxford English Dictionary*. New Haven: Yale University Press.
- Nagy, Andrea R. (2004) Life or Lexicography: How Popular Culture Imitates Dictionaries, *Dictionaries: Journal of the Dictionary Society of North America*, 25: 107-21.
- Nail, Thomas (2017) What is an Assemblage?, *SubStance*, 46(1): 21-37.
- Nardi, Bonnie A. (1995). Studying Context: A Comparison of Activity Theory, Situated Action Models and Distributed Cognition. In Nardi, Bonnie A. (ed.), *Context and Consciousness: Activity Theory and Human-Computer Interaction*, 69-102. Cambridge, MA: The MIT Press.
- Neuhaus, H. Joachim and Marvin Spevack (1975) A Shakespeare Dictionary (SHAD), Some Preliminaries for a Semantic Description, *Computers and the Humanities*, 9: 263-70.
- Newman, William R., ed. (2005). *The Chymistry of Isaac Newton*. <http://webapp1.dlib.indiana.edu/newton/>. Accessed December 22, 2011.
- Nielsen, Jakob (1993) Iterative User-Interface Design, *IEEE Computer*, 26(11): 32-41.

- Nielsen, Sandro and Richard Almind (2011). From Data to Dictionary. In Fuertes Olivera, Pedro and Henning Bergenholtz (eds.), *E-Lexicography: The Internet, Digital Initiatives and Lexicography*, 141-67. London and New York: Continuum.
- Nielsen, Sandro (2011). Function- and User-Related Definitions in Online Dictionaries. In Карташкова, Ф. И. (ed.), *Ивановская лексикографическая школа: традиции и инновации: сб. науч. ст, посвященный юбилею научного руководителя школы, заслуженного работника Высшей школы РФ, доктора филологических наук, профессора Ольги Михайловны Карповой*, 197-219. Иваново: Ивановский Государственный Университет.
- Norberg, Arthur L (1990) High-Technology Calculation in the Early 20th Century: Punched Card Machinery in Business and Government, *Technology and Culture*, 31(4): 753-79.
- Norman, Donald A (1999) Affordance, conventions, and design, *Interactions*, 6(3): 38-43.
- Norman, Donald A. (1988). *The Psychology of Everyday Things*. New York: Basic Books.
- Noys, Benjamin (2000). *Georges Bataille: A Critical Introduction*. London: Pluto Press.
- Nyhan, Julianne (2006). The Application of XML to the historical lexicography of Old, Middle, and Early-Modern Irish: a Lexicon based analysis. National University of Ireland, Cork.
- Nyhan, Julianne (2008) Developing Integrated Editions of Minority Language Dictionaries: The Irish Example, *Literary and Linguistic Computing*, 23(1): 3-12.
- OED *The Oxford English Dictionary. Third Edition. OED Online*. <https://www.oed.com>.
- Okuka, M (1998). *Eine Sprache-viele Erben: Sprachpolitik als Nationalisierungsinstrument in Ex-Jugoslawien*. Wieser Verlag.
- Olney, John, et al. (1968). *Toward the Development of Computational Aids for Obtaining a Formal Semantic Description of English*. Santa Monica: System Development Corporation.

- Olsen, Lance (2000) Architecture of Possibility: Reading the Novels of Milorad Pavić, *The Writer's Chronicle*, 33(1): 34-42.
- Parry, Milman (1971a). *The Making of Homeric Verse: The Collected Papers of Milman Parry*. Oxford: Clarendon Press.
- Parry, Milman (1971b). The Traditional Epithet in Homer. *The Making of Homeric Verse: The Collected Papers of Milman Parry*, 1-190. Oxford: Clarendon Press.
- Patrick, J., et al. (2000) An Architecture and Query Language for a Federation of Heterogeneous Dictionary Databases, *Computers and the Humanities*, 34(4): 393-407.
- Petersen, Marianne Graves, et al. (2004) Aesthetic Interaction: A Pragmatist's Aesthetics of Interactive Systems, Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques: 269-76.
- Pierazzo, Elena (2011) A Rationale of Digital Documentary Editions, *Literary and Linguistic Computing*, 26(4): 463-77.
- Potthoff, Wilfried (ed.) (1990). *Vuk Karadžić im europäischen Kontext*. Heidelberg: Carl Winter Universitätsverlag.
- Jordan, David (ed.) (1994). *Language and Crime in Yugoslavia: Milorad Pavić's Dictionary of the Khazars*. Regionalism Reconsidered: New Approaches to the Field.
- Raymond, Darrell and Frank Tompa (1988) Hypertext and the Oxford English dictionary, *Communications of the ACM*, 31(7): 871-9.
- Rehbein, Malte (2010). The Transition From Classical to Digital Thinking. Reflections on Tim McLoughlin, James Barry and Collaborative Work. In Ryder, Sean and Malte Rehbein (eds.), *Jahresbuch für Computerphilologie*, 55-68. Paderborn: Mentis Verlag.
- Reichmann, Oskar (1989). Lexikographische Einleitung. In Anderson, Robert R., et al. (eds.), *Frühneuhochdeutsches Wörterbuch*, Bd. 1. Einführung, a-äpfelkern, 10-164. Berlin; New York: de Gruyter.

- Reichmann, Oskar (2012). *Historische Lexikographie: Ideen, Verwirklichungen, Reflexionen an Beispielen des Deutschen, Niederländischen und Englischen*. Berlin; Boston: De Gruyter.
- Revard, Carter (1968) On the Computability of Certain Monsters in Noah's Ark: Using Computers to Study Webster's Seventh New Collegiate Dictionary and the New Merriam-Webster Pocket Dictionary, *Proceedings of the 1968 23rd ACM National Conference*, 807-13.
- Revard, Carter (1973) How to Make a N.U.D.E. (New Utopian Dictionary Of English), *Annals of the New York Academy of Sciences*, 211(1): 91-8.
- Revard, Carter (2001a). Interview. <http://amlives.artsci.wustl.edu/transcripts/full/907.pdf>. Accessed October 15, 2014.
- Revard, Carter (2001b). *Winning the Dust Bowl*. Tucson: University of Arizona Press.
- Richards, I. A. (1955). *Speculative instruments*. Chicago: University of Chicago Press.
- Ritchie, Graeme D, et al. (1987) A computational framework for lexical description, *Computational Linguistics*, 13(3-4): 290-307.
- Darnton, Robert (1979). *The Business of Enlightenment: A Publishing History of the Encyclopédie 1775 - 1800*. Cambridge, Mass.: Belknap Press of Harvard Univ. Press.
- Robey, Daniel, et al. (2013) Information Technology, Materiality, and Organizational Change: A Professional Odyssey, *Journal of the Association for Information Systems*, 14(7): 379.
- Robie, Jonathan, et al. (2017). *XML Path Language (XPath) 3.1. W3C Recommendation*. <https://www.w3.org/TR/xpath-3/>.
- Rundell, Michael (1999) Dictionary Use in Production, *International journal of lexicography*, 12(1): 35-53.
- Rundell, Michael (2008). Recent Trends in English Pedagogical Lexicography. In Fontenelle, Thierry (ed.), *Practical Lexicography: A Reader*, 221-43. Oxford, New York: Oxford University Press.

- Russell, Graham J., et al. (1986). A Dictionary and Morphological Analyser for English. *Proceedings of the 11th conference on Computational linguistics*, 277-9. Association for Computational Linguistics.
- Saarela, Mårten Söderblom (2014) Shape and Sound: Organizing Dictionaries in Late Imperial China, *Dictionaries: Journal of the Dictionary Society of North America*, 35(35): 187-208.
- Sager, Naomi (1981). *Natural Language Information Processing: A Computer Grammar of English and Its Applications*. Reading, Mass.: Addison-Wesley.
- Sahle, Patrick (2013). *Digitale Editionsformen - Teil 1: Das typografische Erbe*. Norderstadt: BoD.
- Saussure, Ferdinand de (1916/1981). *Course in General Linguistics*. Glasgow: Fontana Collins.
- Savolainen, Reijo (2016) Elaborating the Conceptual Space of Information-Seeking Phenomena, *Information Research: An International Electronic Journal*, 21(3): not paginated.
- Schrab, Rob (2008) Kangamangus, The Sarah Silverman Program:
- Schreiber, G., et al. (2008) Semantic annotation and search of cultural-heritage collections: The Multimedial E-Culture demonstrator, *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4): 243-9.
- Schreibman, Susan, et al. (2004). *A Companion to Digital Humanities*. Malden, MA: Blackwell Pub.
- Seargeant, Philip (2011) Lexicography as a Philosophy of Language, *Language Sciences*, 33(1): 1-10.
- Seton-Watson, Hugh (1977). *Nations and States: An Enquiry Into the Origins of Nations and the Politics of Nationalism*. London: Methuen.
- Sherman, Donald (1974) A new computer format for Webster's seventh collegiate dictionary, *Computers and the Humanities*, 8(1): 21-6.

- Shillingsburg, Peter (2009) How Literary Works Exist: Convenient Scholarly Editions, *Digital humanities quarterly*, 3: unpaginate.
- Shillingsburg, Peter L. (2006). *From Gutenberg to Google: Electronic Representations of Literary Texts*. New York: Cambridge University Press.
- Siemens, Ray, et al. (2009). *Codex Ultor: Toward a Conceptual and Theoretical Foundation for New Research on Books and Knowledge Environments*. 1(2). http://www.digitalstudies.org/ojs/index.php/digital_studies/article/view/177. Accessed January 30, 2011.
- Sinclair, John (1987). *Looking Up: Account of the Cobuild Project in Lexical Computing*. London: Collins.
- Sinclair, John (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Singleton, David (1999). *Exploring the Second Language Mental Lexicon*. Cambridge; New York: Cambridge University Press.
- Skok, Petar (1971). *Etimologijski rječnik hrvatskoga ili srpskoga jezika*. Zagreb: Jugoslavenska akademija znanosti i umjetnosti.
- Slater, William J (1976) Aristophanes of Byznatium on the Pinakes of Callimachos, *Phoenix*, 30: 234-41.
- Smiljanic, Rajka (2004). *Lexical, Pragmatic, and Positional Effects on Prosody in Two Dialects of Croatian and Serbian : An Acoustic Study*. New York: Routledge.
- Soden, Wolfram von (1994). *The Ancient Orient : An Introduction to the Study of the Ancient Near East*. Grand Rapids, Mich.: W.B. Eerdmans.
- Stankiewicz, Edward (1984). *Grammars and Dictionaries of the Slavic Languages From the Middle Ages Up to 1850: An Annotated Bibliography*. Berlin; New York: Mouton.
- Stendal, Karen, et al. (2016) Analyzing the Concept of Affordances in Information Systems, *System Sciences (HICSS)*, 5270-7.
- Stubbs, M. (1995) Collocations and semantic profiles: on the cause of the trouble with quantitative studies, *Functions of language*, 2(1): 23-55.

- Summers, Della (1988). The Role of Dictioanries in Language Learning. In Carter, Ronald and Michael McCarthy (eds.), *Vocabulary and Language Teaching*, 111-25. London; New York: Longman.
- Svensén, Bo (2009). *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Tarp, Sven (2000) Theoretical Challenges to LSP Lexicography, *Lexikos*, 10: 189-208.
- Tarp, Sven (2008). *Lexicography in the Borderland Between Knowledge and Non-Knowledge: General Lexicographical Theory With Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.
- Tarp, Sven (2011). Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. In Fuertes Olivera, Pedro and Henning Bergenholtz (eds.), *E-Lexicography: The Internet, Digital Initiatives and Lexicography*, 54-70. London and New York: Continuum.
- Tasman, Paul (1957) Literary Data Processing, *IBM Journal of Research and Development*, 1(3): 249-56.
- Tasovac, Toma (2010) Reimagining the Dictionary, or Why Lexicography Needs Digital Humanities, *Digital Humanities 2010*, <http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/html/ab>.
- Tasovac, Toma (2012). Potentials and challenges of WordNet-based pedagogical lexicography: The Transpoetika Dictionary. In Granger, Sylviane and Magali Paquot (eds.), *Electronic Lexicography*, 237-58. Oxford University Press.
- Tasovac, Toma, et al. (2015). Developing Morpho-SLaWS: An API for the Morphosyntactic Annotation of the Serbian Language. *Systems and Frameworks for Computational Morphology. 4th International Workshop, SFCM 2015, Stuttgart, Germany, September 17-18, 2015. Proceedings*, 137-47. Springer.
- Tasovac, Toma (2017). *Libraries in the Age of Digital Infrastructures (and Other Catastrophes)*. https://www.ifla.org/files/assets/academic-and-research-libraries/conferences/toma_tasovac_libraries_dig_infrastructures.pdf. Accessed August 17, 2017.

- Teehan, Aja and John G. Keating (2010a). A Digital Edition of a Spanish 18th Century Account Book: Part 2 – Formalisation and Encoding. In Ryder, Sean and Malte Rehbein (eds.), *Jahrbuch für Computerphilologie*, 189-214. Paderborn: Mentis Verlag.
- Teehan, Aja and John G. Keating (2010b) Appropriate Use Case modeling for humanities documents, *Literary and Linguistic Computing*, 25(4): 381-91.
- Consortium, TEI (2007). *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. <http://www.tei-c.org/Guidelines/P5/>. Accessed 10 May, 2011.
- Tompa, Frank W (1992) An Overview of Waterloo's Database Software for the OED, *Proceedings of the Symposium on Historical Dictionary Databases and Data Retrieval Requirements (Toronto, October 1991)*, 123-43.
- Toms, Elaine (1998) Information Exploration of the Third Kind: The Concept of Chance Encounters, *Information Exploration Workshop for SIGCHI*, 98: <http://www.fxpal.com/chi98ie/submissions/long/toms/index.htm>.
- Toms, Elaine (2000). *Serendipitous Information Retrieval*. Proceedings of the Workshop Information Seeking, Searching, and Querying in Digital Libraries. http://www.ercim.org/publication/ws-proceedings/DelNoe01/3_Toms.pdf. Accessed January 12, 2013.
- Tono, Yukio (2001). *Research on Dictionary Use in the Context of Foreign Language Learning: Focus on Reading Comprehension*. Tübingen: Max Niemeyer Verlag.
- Truesdell, Leon Edgar (1965). *The Development of Punch Card Tabulation in the Bureau of the Census, 1890-1940: With Outlines of Actual Tabulation Programs*. Washington: US Department of Commerce.
- Unsworth, John (2005). *New Methods for Humanities Research. The 2005 Lyman Award Lecture. November 11. National Humanities Center. Research Triangle Park, NC*. <http://www3.isrl.illinois.edu/~unsworth/lyman.htm>. Accessed June 10, 2011.

- Urdang, Laurence (1966) The systems designs and devices used to process the Random House Dictionary of the English Language, *Computers and the Humanities*, 1(2): 31-3.
- Vanhoutte, Edward (2009). Every Reader his own Bibliographer – An Absurdity? In Deegan, Marilyn and Kathryn Sutherland (eds.), *Text Editing, Print and the Digital World*, 99-110. Farnham, England; Burlington, VT: Ashgate.
- Venezky, Richard L (1968) Storage, retrieval, and editing of information for a dictionary, *American Documentation*, 19(1): 71-9.
- Venezky, Richard L (1973) Computer applications In lexicography, *Annals of the New York Academy of Sciences*, 211(1): 287-92.
- Verkuyl, Henk, et al. (2003). The Codification of Usage by Labels. In Sterkenburg, Piet van (ed.), *A Practical Guide to Lexicography*, 2003, 297-311. Amsterdam/Philadelphia: John Benjamins.
- Verlinde, Serge and Jean Binon (2009). Pedagogical Lexicography Revisited. In Bergenholtz, Henning, et al. (eds.), *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*, 69-90. Bern: Peter Lang.
- Verlinde, Serge (2010). The Base lexicale du français: a multi-purpose lexicographic tool. In Granger, Sylviane and Magali Paquot (eds.), *eLexicography in the 21st Century: New Challenges, New Applications*, 335-42. Louvain-la-Neuve: UCL.
- Vitalich, Kristin Leigh (2005). Lexicographical doxa: The writing of Slavic dictionaries in the nineteenth century (Samuel Bogumil Linde, Vuk Stefanović Karadžić, Vladimir Ivanovich Dal). University of California at Los Angeles.
- Wachtel, Andrew (1997) Postmodernism as Nightmare: Milorad Pavic's Literary Demolition of Yugoslavia, *The Slavic and East European Journal*, 41(4): 627-44.
- Walker, Donald E (1986) Knowledge Resource Tools for Accessing Large Text Files, *Information in Data: Proceedings of the First Conference of the University of Waterloo Center for the New Oxford English Dictionary*, 1-22.

- Walker, Donald E (1989) Developing Lexical Resources, *Dictionaries in the Electronic Age: Proceedings of the 5th Annual Conference of the UW Centre for the New Oxford English Dictionary*, 1-22.
- Walmsley, Priscilla (2015). *XQuery : Search Across a Variety of XML Data*. Sebastopol, CA: O'Reilly Media.
- Webster, Noah (1789/1991). An Essay on the Necessity, Advantages and Practicability of Reforming the Mode of Spelling, and of Rendering the Orthography of Words Correspondent to the Pronunciation. In Crowley, Tony (ed.), *Proper English?: Readings in Language, History, and Cultural Identity*, Politics of language, London; New York: Routledge.
- Wells, Ronald A. (1973). *Dictionaries and the Authoritarian Tradition: A Study in English Usage and Lexicography*. The Hague: Mouton.
- Wermser, Richard (1976). *Statistische Studien zur Entwicklung des englischen Wortschatzes*. Bern: Francke.
- Wiegand, Herbert Ernst (1989). Der gegenwärtige Status der Lexikographie und ihr Verhältnis zu anderen Disziplinen. In Hausmann, Franz Josef, et al. (eds.), *Wörterbücher: ein internationales Handbuch zur Lexikographie*, 1, 246-80. Berlin; New York: W. de Gruyter.
- Wiegand, Herbert Ernst (1998). *Wörterbuchforschung: Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. Berlin: W. de Gruyter.
- Wiegand, Herbert Ernst (1994/2000). Über die Mediostrukturen bei gedruckten Wörterbüchern. In Kammerer, Matthias (ed.), *Kleine Schriften: Eine Auswahl aus den Jahren 1970 bis 1999 in zwei Bänden*, 1, 1163-92. Berlin; New York: DeGruyter.
- Wiegand, Herbert Ernst (2000). *Kleine Schriften: Eine Auswahl aus den Jahren 1970-1999 in zwei Bänden. Bd 1: 1970-1988*. Walter de Gruyter.

- Wiegand, Herbert Ernst (2001) Was eigentlich sind Wörterbuchfunktionen?
Kritische Anmerkungen zur neueren und neusten Wörterbuchforschung,
Lexicographica, 17: 168-252.
- Wierzbicka, Anna (1992). *Semantics, Culture and Cognition: Universal Human Concepts in Culture-Specific Configurations*. New York: Oxford University Press.
- Wilks, Yorick, et al. (1996). *Electric Words: Dictionaries, Computers, and Meanings*.
Cambridge, MA: MIT Press.
- Williams, John N (1992) Processing Polysemous Words in Context: Evidence for
Interrelated Meanings, *Journal of Psycholinguistic Research*, 21(3): 193-218.
- Wilson, D (1986). *The life and times of Vuk Stefanović Karadžić, 1787-1864: Literacy, Literature, and National Independence in Serbia*. Oxford: Clarendon Press.
- Wilson, Tom D. (1997) Information behaviour: an interdisciplinary perspective,
Information processing & management, 33(4): 551-72.
- Wölck, Wolfgang (1969) A Computerized Dictionary of Andean Languages, *Language Sciences*, 8: 1-9.
- Yong, Heming. and Jing. Peng (2008). *Chinese lexicography: A History From 1046 BC to AD 1911*. Oxford; New York: Oxford University Press.
- Young, Robert J. (1996). *France and the Origins of the Second World War*. New York: St. Martin's Press.
- Zampolli, Antonio (1973) Humanities Computing in Italy, *Computers and the Humanities*, 7(6): 343-60.
- Zečević, Vesna (2000). *Hrvatski dijalekti u kontaktu*. Zagreb: Institut za hrvatski jezik i jezikoslovlje.
- Zgusta, Ladislav (1972). *Manual of Lexicography*. The Hague: Mouton.

- Zgusta, Ladislav (1989). The Role of Dictionaries in the Genesis and Development of the Standard. In Hausmann, Franz Josef, et al. (eds.), *Wörterbücher: ein internationales Handbuch zur Lexikographie*, 1, 70-9. Berlin; New York: W. de Gruyter.
- Албијанић, Александар (2010). *Прилози о српском предвуковском књижевном језику*. Београд: Завод за уџбенике.
- Аперсян, Ю. Д. (1988). Прагматическая информация для толкового словаря. In Арутюнова, Нина Давидовна (ed.), *Прагматика и проблемы интенциональности (Сб. науч. тр.)*, Москва: Институт языкознания АН СССР.
- Белић, Александар and Петар Пијановић (1999). *Дијалекти источне и јужне србије*. Београд: Завод за уџбенике и наставна средства.
- Белић, Александар (2000). *О дијалектима*. Београд: Завод за уџбенике и наставна средства.
- Бјелетић, Марта, et al. (2003). *Етимолошки речник српског језика*. Београд: Институт за српски језик САНУ.
- Вујаклија, Милан (1937/2004). *Лексикон страних речи и израза*. Београд: Просвета.
- Вукомановић, Славко (1976) Тумачење речи у Вуковом Српском рјечнику од 1852. године, *Научни састанак слависта у Вукове дане*, 5: 705-17.
- Вукомановић, Славко (1999) Речи и њихове одреднице у Вуковом другом српском рјечнику, *Српски језик*, 4(1/2): 203-25.
- Грицкат, Ирена (1966) У чему је значај и какве су специфичности славеносрпског периода у развоју српскохрватског језика: поводом иницијативе Матице српске да се састави речник књижевног језика славеносрпског периода, *Зборник за филологију и лингвистику*, 9: 61-6.
- Деретић, Јован. (1990). *Кратка историја српске књижевности*. Београд: БИГЗ.
- Дешић, Милорад (1996) Дефиниције у Вуковом „Српском рјечнику” из 1852. године, *Научни састанак слависта у Вукове дане*, 25(2): 217-22.

- Дмитриев, Петр Андреевич and Герман Иванович Сафронов (1984). *Вук С. Караджич и его реформа сербскохорватского/хорватосербского литературного языка*. Ленинград: ЛГУ.
- Добрашиновић, Голуб, et al. (1980). *Копитар и Вук*. Тршић и Београд: Вуков сабор; Рад.
- Елаковић-Ненадовић, Ана (2011) Књижевни језик као језик српске културе у цивилизације: Вук Караџић и Валтазар Богишић на заједничком послу, *Научни састанк слависта у Вукове дане*, 40(1): 567-76.
- Елезовић, Глигорије (1932). *Речник косовско-метохиског дијалекта*. Београд: Српска краљевска академија.
- Елезовић, Глигорије (1935). *Речник косовско-метохиског дијалекта*. Београд: Српска краљевска академија.
- Засорина, Лидия Николаевна (1966). *Автоматизација и статистика в лексикографији: работа над частотным словарем русского языка*. Ленинград: Издательство Ленинградского университета.
- Златановић, Момчило (1997). *Речник говора Јужне Србије : (провинцијализми, дијалектизми, варваризми и др.)*. Врање: Учитељски факултет.
- Иванова, Најда И. (2010). *Славеносрпски језик између “простоте” и “совершенства”*.
- Ивић, Милка (1988) О Вуковој лексици и поводом ње, *Научни састанак слависта у Вукове дане*, 17(1): 23-8.
- Ивић, Милка (1990). *О језику Вуковом и вуковском*. Нови Сад: Књиж. заједница Новог Сада.
- Ивић, Милка (2003) Кулинарски рецепти у Вуковом рјечнику, *Даница*, 10: 318-22.
- Ивић, Павле (1966). *О Вуковом Рјечнику из 1818. године*. Београд: Просвета.
- Ивић, Павле (1971). *Српски народ и његов језик*. Београд: Српска књижевна задруга.

Ивић, Павле (2009). *Српски дијалекти и њихова класификација*. Сремски Карловци: Стојановић.

Караџић, Вук Стефановић (1818). *Српски рјечник, истолкован њемачким и латинским ријечима*. *Serbisch-Deutsch-Lateinisches Wörterbuch. Lexicon Serbico-Germanico-Latinum*. Беч: Јерменска штампарија.

Караџић, Вук Стефановић (1852a). Предговор. *Српски рјечник истумачен њемачкијем и латинскијем ријечима*, Беч: Јерменска штампарија.

Караџић, Вук Стефановић (1852b). *Српски рјечник истумачен њемачкијем и латинскијем ријечима*. *Lexicon Serbico-Germanico-Latinum*. Беч: Штампарија јерменскога намастира.

Караџић, Вук Стефановић (1892). *Српски рјечник истумачен њемачкијем и латинскијем ријечима*. *Треће издање државно*. *Lexicon Serbico-Germanico-Latinum. Editio tertia publica*. I. Edited by Јован Бошковић. Београд: Штампарија Краљевине Србије.

Караџић, Вук Стефановић (1898). *Српски рјечник истумачен њемачкијем и латинскијем ријечима*. *Треће (државно) издање, исправљено и умножено*. *Lexicon Serbico-Germanico-Latinum. Editio tertia (publica), emendata*. Edited by Пера Ђорђевић and Љубомир Стојановић. Београд: Штампарија Краљевине Србије.

Караџић, Вук Стефановић (1935). *Српски рјечник истумачен њемачкијем и латинскијем ријечима*. *Четврто државно издање*. *Lexicon Serbico-Germanico-Latinum*. Београд: Штампарија Краљевине Србије.

Караџић, Вук Стефановић (1816/1969). *Објавленіе о Србскоме Рѣчнику*. In Ивић, Павле (ed.), *Српски рјечник (1818)*, 3-9. Београд: Просвета.

Караџић, Вук Стефановић (1852/1986a). *Српски рјечник (1852)*. I. Edited by Јован Кашић. Београд: Просвета.

Караџић, Вук Стефановић (1852/1986b). *Српски рјечник (1852)*. II. Edited by Јован Кашић. Београд: Просвета.

- Кашић, Јован (1973) Фонолошки, морфолошки и семантички дублети у Вуковом рјечнику, *Годишњак Филозофског факултета*, 15(1): 105-42.
- Константиновић, Радомир (1964). Вук расковничар. In Константиновић, Радомир (ed.), *Расковник: проза из рјечника*, 7-20. Београд: Просвета.
- Кречмер, Ана (1988) Српски књижевни језик у деценијама пре Вука (језичка анализа славеносрпских текстова: проблеми и могућности), *Научни састанак слависта у Вукове дане*, 17/2: 139-49.
- Кречмер, Ана (2007). Славеносрпска писменост и њено значење за историјску србистику. *Шездесет година института за српски језик САНУ. Зборник радова I*, 313-20. Београд: Институт за српски језик САНУ.
- Кулаковский, Платон Андреевич (2005). Вук Караджич. *Его деятельность и значение в сербской литературе*. Москва: УРСС.
- Латковић, Видо (1964) Примери из народне епике у другом издању Вукова Рјечника, *Анали Филолошког факултета*, 4: 243-8.
- Матијашевић, Јелка (1988). Вуков Рјечник као извор грађе за социолингвистичка проучавања. Београд: МСЦ.
- Матић, Светозар (1960) Народна песма у Вуковом “Рјечнику”, *Зборник Матице српске за књижевност и језик*, 8: 139-45.
- Матицки, Миодраг (1978). Вишњићеве песме као грађа Вуковог Српског рјечника. *Ковчежић: прилози и грађа о Доситеју и Вуку*; књ. 16, 46-52. Београд: Вуков и Доситејев музеј.
- Медведев, Никола (1990) Лична имена “Српског рјечника”, *Зборник радова Техничког факултета “Михајло Пупин”*, 1(1): 108-18.
- Милинчевић, Васо (1987) Вуков “Српски рјечник” као научно и књижевно дело, *Књижевна историја*, 20(75-78): 83-90.
- Младеновић, Александар (1989). *Славеносрпски језик: студије и чланци*. Нови Сад: Књижевна заједница Новог Сада.

- Младеновић, Живомир (1966) Рукописни извори одломака народних песама у Вуковом “Српском рјечнику”, *Анали Филолошког факултета*, 4: 281-92.
- Москович, В. А. (1966) Автоматизација неких аспеката лингвистичког рада, *Вопросы языкознания*, XV(1): 102-10.
- МС (1967). *Речник српскохрватског књижевног језика*. Нови Сад: Матица српска.
- Павић, Милорад (1984). *Хазарски речник: роман лексикон у 100,000 речи*. Београд: Просвета.
- Папп, Ференц (1968) О плану машинне обраде етимолошког речника венгерског језика, *Советское финно-угроведение*, 3: 207-14.
- Пецо, Асим (1986) Један детаљ из проблематике: турцизми у Вуковим рјечницима, *Књижевни језик*, 15(3/4): 305-9.
- Пецо, Асим (1987а) Германизми у Вуковом рјечнику из 1818. године, *Трећи програм*, 15(55): 318-24.
- Пецо, Асим (1987б). *Турцизми у Вуковим рјечницима*. Београд: БИГЗ.
- Поповић, Миодраг (1983). *Памтивек: Српски рјечник Вука Ст. Караџића*. Београд: Завод за уџбенике и наставна средства.
- САНУ (1959). *Речник српскохрватског књижевног и народног језика*. Београд: Институт за српскохрватски језик.
- Селимовић, Меша (1967). *За и против Вука*. Нови Сад: Матица српска.
- Симић, Радоје (1981). Вук-Даничићева акценатска система и питање акценатских парадигми. In Исаковић, Антоније and Јосип Торбарина (eds.), *Зборник о Ђури Даничићу*, 251-7. Београд: Српска академија наука и уметности.
- Симић, Радоје (1988) Стилистичка стратификација језика у Вукову “Српском рјечнику”, *Научни састанак слависта у Вукове дане*, 17/1: 215-26.
- Скерлић, Јован (1966). *Српска књижевност у XVIII веку*. Београд: Просвета.

- Слапшак, Светлана (1987). *Вуков Рјечник и преведенице са грчког*. Нови Сад: Књижевна заједница Новог Сада.
- Спремић, Момчило. (1994). *Деспот Ђурађ Бранковић и његово доба*. Београд: Српска књижевна задруга.
- Стакић, Милан В. (1988) О пејоративима у Вуковом Рјечнику, *Научни састанак слависта у Вукове дане*, 17/1: 153-9.
- Станић, Милија (1978). О неким дефиницијама у Вукову Српском рјечнику. *Ковчежић: прилози и грађа о Доситеју и Вуку*; књ. 16, 80-4. Београд: Вуков и Доситејев музеј.
- Стојановић, Љубомир (1987). *Живот и рад Вука Стеф. Караџића*. Београд: Београдски издавачко-графички завод.
- Сувајдић, Бошко (2011) Вук и предвуковиски записи, *Научни састанак слависта у Вукове дане*, 40/2: 201-14.
- Тасовац, Тома (2007). *О култури речника*. <http://www.politika.rs/rubrike/Kultura/t44709.sr.html>.
- Тасовац, Тома and Ивана Башић (2012) Корпусна прагматика као изазов лексикографској пракси, *Зборник Матице српске за филологију и лингвистику*, LV(2): 39-52.
- Тјапко, Галина Георгијевна (2011) Развој граматичких погледа Вука Караџића у његовој лексикографској пракси, *Научни састанак слависта у Вукове дане*, 40(1): 341-50.
- Фрумкина, Ревекка Марковна (1964) Автоматизация исследовательских работ в лексикологии и лексикографии, *Вопросы языкознания*, XII(2): 114-9.
- Цветковић, Ирена (2005) Славенизми у Вуковом Српском рјечнику из 1852. године, *Зборник Матице српске за филологију и лингвистику*, 48(1-2): 95-108.
- Леовац, Славко (ed.) (1987). *Вуков лексикографски поступак (на примеру односа према 'вулгаризмима' и хомонимији*. Зборник радова о Вуку Караџићу. Сарајево: Институт за књижевност.

- Штасни, Гордана Р. (2012) Етници у Вуковом Српском рјечнику (1852) и лексичко-семантичка норма, *Научни састанк слависта у Вукове дане*, 41/1: 309-19.
- Щерба, Лев Владимирович (1940) Опыт общей теории лексикографии, *Известия Академии наук Союза СССР, Отделение литературы и языка*, 3: 89-117.