Broadening Perspectives in the History of Dictionaries and Word Studies

Edited by

Hans Van de Velde

Fredric T. Dolezal

Broadening Perspectives in the History of Dictionaries and Word Studies

Broadening Perspectives in the History of Dictionaries and Word Studies

Edited by

Hans Van de Velde and Fredric T. Dolezal

Cambridge Scholars Publishing

Broadening Perspectives in the History of Dictionaries and Word Studies

Edited by Hans Van de Velde and Fredric T. Dolezal

This book first published 2021

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-7432-6
ISBN (13): 978-1-5275-7432-8

TABLE OF CONTENTS

Introduction

Broadening Perspectives in the History of Dictionaries and Word Studies

Fredric T. Dolezal and Hans Van de Velde

> *It goes without saying that history$_{[narration]}$ is possible only because there are traces of history$_{[events]}$, usually called "sources"; and it also stands to reason that old dictionaries, being themselves part of history$_{[events]}$ are eminent sources for history$_{[narration]}$, or historiography of language. (Zgusta 2006, 5)*
>
> *Dem Wörterbuch eignet das Wort dann erst, wenn es mit einem Begriff inning verbunden und in gemachter Form in irgend einnem Sprachdenkmal erscheint.*
> *[The word belongs in the dictionary only when it is intimately connected with a notion and when having an established form it occurs in a text.] (Passow 1812, 8; [Zgusta 2006, 27])*[1]

1. Words, dictionaries, and the literariness of lexicography

Franz Passow claimed that every word has a history, and to this one might add that every word with a history also has a bibliography. From the perspective of literary lexicography, a history of dictionaries becomes an exploration of the bibliographic repository of literary history; by combining the critical literary and linguistic approach the policies and decisions of the lexicographer are embedded in cultural and aesthetic domains. Dictionaries and the word studies found within them are reflections not only of a

[1] (1) *Lexicography Then and Now: Selected Essays*. Zgusta, Ladislav. 2006. Eds. Fredric S.F. Dolezal and Thomas B.I. Creamer. Max Niemeyer Verlag: Tübingen; (2) Passow, Franz. 1812. *Über Zweck, Anlage und Ergänzung Griechischer Wörterbücher*. Berlin: F. Maurer.

linguistic community in the narrowest conception, but also a repository and reflection of the culture and literary communities of the time in which the dictionaries are published. However, print or virtual dictionaries, as all bibliographic texts, can only be incomplete repositories; broadly speaking, some texts and words have been lost over time, just as some words have never been collected from texts that do still exist.

The research found in the following pages of this volume continues developing and expanding previous studies in the series of volumes dedicated to historiographical lexicography and lexicology (see Section 3 below for a list of previous *International Conference for Historical Lexicography and Lexicology* meetings and publications)—for instance, the study of manuscripts and print artifacts; dictionaries and standard varieties; biographies; bibliography and text analyses; dictionary production; a variety of critical literary and linguistic perspectives; and corpus and digital analyses. The proof and demonstration of the joint literary and linguistic basis of lexicography and lexicology is well inscribed in the following pages of this volume.

Broadening Perspectives in the History of Dictionaries and Word Studies represents a selection of papers originally presented at the *10th International Conference on Historical Lexicography and Lexicology* (*ICHLL*) held on June 12-14, 2019 at the Fryske Akademy at Leeuwarden (the Netherlands). The theme of ICHLL 10 was 'The history of lexicography and lexicology of smaller languages' with papers on other topics in the history of lexicography and lexicology also welcomed. Each ICHLL provides a platform for scholars and researchers to present their papers and share their research, their experiences, their ideas, and the practical challenges encountered and/or the solutions adopted concerning the history of dictionaries, the making of historical dictionaries, as well as historical lexicology. Thirty-seven papers were originally presented at the conference. The chapters in this book are expanded and revised, peer-reviewed versions of a selection of those papers that were submitted for publication.

ICHLL 10, proposed by Anne Dykstra, was locally sponsored by the *Fryske Akademy*. The organizers and conveners of the conference were Hans Van de Velde (chair, Fryske Akademy), Rolf H. Bremmer Jr (Leiden University/Fryske Akademy), and Willem Visser (secretary, Fryske Akademy/Rijksuniversiteit Groningen). Since its foundation in 1938, Fryske Akademy has been the lexicographic institute for Frisian.

Each of the meetings of ICHLL is locally sponsored under the aegis of the *International Society for Historical Lexicography and Lexicology* (ISHLL) and embraces the broad range of topics and disciplinary approaches of the scholarly interests of the members of ISHLL[2].

2. An Overview

The first contribution, from **Chiara Bertulessi**, opens up a discussion of the relationships between lexicography and ideology ("a specific representation of reality within a given context") from the theoretical perspective of critical lexicography. She defines this analytical approach as a "research field that falls within the scope of metalexicography". This article then adds to the growing attention to the development and definition of critical lexicography, and consequently challenges a greater scholarly attention to metalexicography as part of, not distinct from, lexicography as practice.

This interplay of production and reception is framed in the context of how the leadership of the People's Republic of China has placed importance on lexicographic theory, practice and production within the nation. The emphasis on language reform and lexicographical activities began soon after the founding of the People's Republic in 1949. Bertulessi offers the reader a "diachronic perspective to identify the changes in the ways in which the political objectives associated to the field of lexicography have changed since the mid-1970s". More generally, looking at a wider range of dictionary production and history, Bertulessi calls for a perspective that addresses how the production of dictionaries, and all the related activities that entails, support the construction and reception of dominant ideologies within a language community. She performs the specific critical study of lexicography in the People's Republic of China through a close reading of official documents that show an exchange and contestation of linguistic and lexicographic ideas that guide an assessment of dictionary history and the progress of dictionary production.

Other articles in this collection also ask questions about the relationships between "reform", "standardization", and the "prescription of linguistic standards" in the respective vernaculars; we shall see that these notions that seem always present, covertly or overtly, within discussions of grammars and dictionaries deserve a wider perspective than the conventional binary description/prescription formula supplies.

[2] ISHLL is a society of scholars working on the world history of dictionaries, the making of all types of dictionaries across languages and cultures, critical lexicography, and historical lexicology. It was established as a result of the first *International Conference on Historical Lexicography and Lexicology*, which was organized by Julie Coleman and took place in Leicester, England in 2002. For further information: http://calepino.uga.edu/wordpress/

Bertulessi addresses the question by directly placing standards, prescription and common language within a context of "class character" as a reflection of ideological and political necessity. In this way, sociolinguistic and cultural realities are put in the foreground side by side with compiling dictionaries, that show "innovations and quality" which meet the demands of internationally recognized standards of linguistics scholarship, lexicography and lexicology. All of this history comes with passionate debate and political costs and benefits, which Bertulessi documents with clarity.

The end result of this dialectic in the recent history of Chinese lexicography, especially as it produces learners' dictionaries for speakers from diverse linguistic communities, aims for the "cultural soft power" of becoming a world-class "lexicographic power". And, as an instance of one history, she also stresses the general importance for investigating and describing the broader cultural contexts as a topic across all traditions and histories that critically reveal past and current practice by the close reading of lexicographic artifacts.

The next article, by **Harald Bichlmeier and Güler Doğan Averbek**, is a description and analysis of the Turkish-German versified dictionary titled *Almanca Tuhfe / Deutsches Geschenk* [German Gift], published by Dr. Şerefeddin Pasha in 1916. They make a case for versified glossaries having relevance and value in the research domains of lexicology, pedagogical scholarship and children's literature. They are particularly interested in the *Almanca Tuhfe*'s place in the history of bilingual and multilingual glossaries written in verse for the purpose of teaching language learners core vocabulary. These glossaries, or simplified dictionaries, were prevalent in Islamic cultures in order to teach Arabic. The pragmatic purpose of using poetic form as a mnemonic device for learning has been a standard in the history of learning and teaching from the earliest documented times. A comprehensive history of dictionaries reveals their central place in education theory and practice.

The *Almanca Tufhe* follows a well-known formula for writing versified multi- and bilingual dictionaries in the Islamic tradition. A significant feature of the macrostructure features the poet (in this case also the lexicographer) addressing topics about textual sources consulted, a critique of previous versified glossaries, followed by stating the purpose for the writing of the glossary at hand. All of this within a poetic meter and stanzaic form that is familiar to readers of these dictionaries. As Bichlmeier and Doğan Averbek show, Dr. Şerefeddin Pasha did not feel constrained by the poetic tradition. The author did employ the practice of the *aljamiado* literature, that is, the use of Arabic script to write out the target language; in this case, the German words are rendered in Arabic script.

The essential purpose for such a glossary is the result of the cultural and political exchange between the Ottoman and German states. Because Germany sent “missionaries” representing professions such as physicians, clerics, teachers, and scholars and promoted the opening of schools, a knowledge of German was considered of prime importance according to Şerefeddin Pasha because of the growing exchange between Germany and the Ottoman Empire. The use of Arabic script in 1916 for those who are familiar with the founding of modern Turkey by Kemal Ataturk and his romanization of Turkish writing put a time stamp on the usefulness of Arabic script. Bichlmeier and Doğan Averbek through their continuing descriptive and analytical readings of this text offer us an example of how language contact and its consequences on people and cultures can be traced by studying this tradition of versified glossaries.

As we turn to the third selection, **Andreas Deutsch** describes the ongoing work on the *Deutsches Rechtswörterbuch* founded in 1897. Along with his own wide-ranging project on German legal language and related topics, here he also develops the continuing historiographical narrative of Jakob Grimm’s excavation of German language and culture, in this instance, his work on “German Legal Antiquities” (*Deutsche Rechtsalterthümer* 1828). “German” in this telling includes for Grimm languages of the West Germanic language family, Frisian, Dutch, Old Saxon and Anglo-Saxon; and it is in Old Frisian that an essential historical record of German legal terms can be found, because Old Frisian language has been primarily preserved in legal sources. This article among other items of note shows how these sources of technical vocabulary “provide an exciting insight into past living conditions.”

Here we see how a critical historical lexicography can also be described as an instance of literary lexicography. As Deutsch points out Grimm’s dictum: “The law book of the old Frisians contains throughout its oldest and best part pure and precious poetry.” And to make the point clearer attention is directed to “the common roots of law and literature”, or in the words of Grimm (see footnote 2 in the article for the original text), “It is not difficult to believe that law and poetry rose from the same bed.”

It is the production of the *Deutsches Rechtswörterbuch* that collects and presents evidence not only for the technical language of the law, but also those linguistic items of everyday language that had legal relevance. Deutsch tells us that at least 1000 new entries will be added each year to the current editions of the legal dictionary, which is expected to encompass 16 volumes and around 100,000 entries. In this present article, the focus is on the special place of Old Frisian from the perspective of historical research on legal terminology and its parallel relevant lexicographic histories.

The record of another language which in its older form can be found mostly in legal documents is Faroese. **Chiara Benati** brings our attention to the people and circumstances responsible for the survival of the language. An essential and critical person in this history is Jens Christian Svabo (1746-1824). His work on finding and collecting ballads became foundational for the progress of Faroese lexicography. Svabo compiled the first dictionary of the language, *Dictionarium Faeroense,* which was not published for about 200 years after he produced the manuscript. In this contribution to our widening appreciation of a more global historical lexicography, Benati emphasizes the close relationship between the preservation of the poetic language of the ballads—as a trace of the language in other contexts—and the ability to later construct historical and linguistic basis for establishing a standard for Faroese. She points out that up to this time there has been scant attention to Svabo's work in lexicography and collections of the oral traditions of the people, a circumstance that this present study begins to remedy with her close reading and analysis of the relevant artifacts in manuscript and print.

As with other articles in this collection we are asked to consider the relationships among the emancipation and recognition of a linguistic community (the right and ability to speak and write their own language), the literary traditions of the community and the impulse to establish a standard language. At times this work may seem to be (or actually be) the prescriptive dictate of the producers of dictionaries; and, at other times, it is the effort to bring clarity, mutual understanding and the recognition of a linguistic community that underlies the importance of compiling dictionaries. Benati describes the historical situation of Faroese, with the competing interests of the Danish language alongside the minority Faroese language, as a "a bunch of spoken dialects without any standard norm, while Danish was the official language, used in writing and in every official circumstance." The simplistic commonplace binary of prescription and/vs. description must give way to a wider linguistic and cultural perspective.

Standard, substandard, corruption of language and marginal category have the sound of earlier epithets of language theorists and critics. And yet, as **Rachel Fletcher** through a careful critical analysis of dictionary texts shows, these terms and their consequences are the keywords in the contested periodization of Old English, namely the use over time of "late Old English" that still has relevance for deciding which words to include in "standard" Old English. The considerations of language contact, dialect and register are important linguistic factors in the development of the lexicographic tradition of Old English.

Fletcher employs an historical perspective through the study of seven major dictionaries of Old English (seventeenth century to the present day) to describe “the creation and handling of late Old English as a marginal category within a wider lexicographical tradition”. Fletcher makes the point that the liminal character of lateness in the Old English period contributed to the sub-period receiving less scholarly attention and misrepresenting its language and literature, thereby confusing rather than understanding a more neutral description of the period.

In this case, “corruption of the standard” applies to the scholarly and historical assessment of language use rather than a document of attitudes of the language users themselves. The poetic is also of importance here. Fletcher shows that at times the issue of marginality is not a matter of when but a matter of where the linguistic item occurs in the sources available: that is, if an item is only found in the register of poetry. All of this is the result of discrete comparison of a series of dictionaries and their makers. Much of what she finds is the result of a nuanced and close reading of dictionaries as texts, and an exploration of how ideas of periodization of English continue as a point of debate in English lexicography, its theory and practice and historical linguistics. Fletcher emphasizes the variety of descriptive techniques to show the subordinate status of late Old English in this history and how the items being considered are not analyzed synchronically, nor in relation to the period of English that follows, namely Middle English.

As **Volker Harm** tells us in his critical history of Grimmian lexicography and its inheritors, German lexicography has neither a comprehensive national dictionary project, nor “a common reference point within the lexicographic landscape.” In this article, Harm provides a retrospective account of the DWB2, a dictionary that has not garnered the critical attention that one might reasonably consider not only appropriate but essential in the history of German lexicography in particular, and historical lexicography in general. The monumental task of revising the first edition of the DWB can be shown by a simple comparison of citation slips collected for A-Z of the first edition (2.5 million) and the number for A-F of the second (5.5 million). With the publication of the last fascicle of the second edition of the *Deutsches Wörterbuch*, the *Neubearbeitung des Deutschen Wörterbuch,* completed as of 2018 (180 years after the Grimms began their project), broadening perspectives of historical dictionary practice has become a lexicographical imperative.

Sufficient resources for the work to continue with the same quality of scholarship are not available, especially considering the timeline that would have to be supported. Even the Online (Third) Edition of the *Oxford English Dictionary* is undergoing a consideration of how much of their resources can be given to enhancing and raising the quality of older entries. Harm asks if it is possible to continue historical lexicography in the Grimmian style; that is, with the Passowian emphasis on a positivistic narrative based on each word telling its own story through an accumulation of textual sourcing. And thus, Harm further asks, "Are completely new ways of doing historical lexicography imaginable", and, can a new way entail preserving a dictionary in the tradition of the original that has a strong narrative component that connects the present through a documented evolution of meanings through 400 years of German literature and language?

As he lays out in the second major section of his article, the answer is fortunately, yes, there is a way, and it comes in the form of a dictionary, called *Wortgeschichte digital*. The lexicographers intend to produce a dictionary as the Grimms intended that will "make the linguistic past interesting for the audience of their time". The research team has chosen an onomasiological structure so that with "the words belonging to a common field … the user gets a far more coherent picture of the lexical developments". They also hope to appeal to a variety of users, those who are patient readers of the discursive sections of an entry, those who are more interested in finding a word and its history or meaning and those who may want to perform a database search across words and ideas. Harm believes that all these modification and enhancements would have met with the approval of the Grimms, who themselves continually made revisions to their own lexicographical procedures and methods; or as Jakob Grimm once wrote in a letter: "bücher dieser art können erst gut werden bei zweiter auflage" ["books of this kind can only become real good ones with their second edition".]

"Old Onomatopoeia: What Etymological Dictionaries Tell us About Sound Imitation in Extinct Languages" by **Maria Flaksman** opens up a well-sourced and discretely descriptive and explanatory discussion on etymological dictionaries as a source for information about the historical development of phonetically imitative words (English *plop*!, or Russian *жужжать* 'to buzz'). She points out that identifying "the imitative origin of a word" is a primary concern for establishing a diachronic approach to iconicity studies. Flaksman provides an historical overview of the terminology that can be found in the literature going back to Charles Peirce ('iconic'), through Ferdinand de Saussure ('symbolic') up to the present day ('phono-symbolic'; 'sound symbolic'; 'expressives'). The other concern addressed in this article covers the question of whether an imitative system remains stable over the course of its evolution.

Flaksman clearly shows how historical-comparative research and etymological studies are important for current studies in phonosemantic analysis. The methodology she engages relies on current linguistic methods and practice combined with a wide-ranging and resource-rich diachronic perspective of sound change. Her research has found that imitative words can lose the "iconic link", that is the relations of form and meaning, as a language changes semantically and by way of regular sound changes over time. The process of newer expressive words, sometime replacing, sometimes adding to the existing lexicon of imitative sounds, makes the diachronic analysis difficult to assess even with access to an etymological record. In other words, "regular" patterns of sound change can obliterate the obvious iconicity of a word and leave it with the well-known linguistic attribute of the arbitrariness of sound-meaning. Fortunately, Flaksman has found that etymological dictionaries are good resources that provide evidence for establishing phonemic inventories. Her acutely developed quantitative and qualitative analysis of dictionaries over time and the use of current linguistic methodologies broaden our perspective of a distinctive feature of language change.

With the availability of the Royal Society Corpus with its finely tuned metadata that comes with 48,000 digitized texts from the early days of the Royal Society to the present day, we have new and enhanced possibilities for well-motivated database searches in tandem with qualitative linguistic analyses to support historical research in lexicography and lexicology. **Katrin Menzel** shows one such intriguing possibility with clear implications for others as she presents the significant outcomes of textual study and prospects for future lexicographical production in her chapter on "Scientific Eponyms throughout the History of English Scholarly Journal Articles". Existing reference works that treat eponymous words can be complemented with a more systematic exploration of not only historical and present-day trends in eponymization, but also a critical understanding of why particular eponyms were coined and how they spread within the scientific community.

With the availability of digitized scientific articles covering a few hundred years, using textually based analysis and explication Menzel provides insight into cultural and scientific developments that affect meanings of eponyms; she also shows how a balance of quantitative and qualitative resource-based analysis can underwrite lexicological investigations that consider the "literal" and "metaphorical" and the establishment of scientific consensus. Menzel believes that the use and addition of "natural and authentic" illustrative quotations will broaden the usefulness of the information available in existing dictionaries of eponyms, or entries treating eponyms in non-specialized dictionaries. A dictionary user interested in eponyms will have a wider perspective with more and clearer access to multiword expressions and frequent or significant collocations.

In the course of her article, Menzel also considers the formation of scientific consensus, the identification of regularities of linguistic change and word formation, lexical ambiguities (one proper name appearing in more than one term), the contribution to translation studies, the male dominant naming choices and current trends in eponym formation, among other insights that lexicographers and historians of science will find useful, even necessary in their own methods and practices. It is still the case, notwithstanding a move away from the use and formation of eponyms, that the impulse to eponymize can still be found in new articles across the scientific disciplines covered in the Royal Society Corpus.

The next chapter also emphasizes a resource-based lexicological study that makes significant contributions to a widening study of dictionary history and historical lexicography. **Petra Storjohann** develops her historiographical and linguistic analysis in "German Paronym Dictionaries: The Shift from Prescriptive Print Editions to Electronic Corpus-based Resources." The study of paronyms, or "confusables", explores the sort of words that people mix up with other words that are brought to mind because of their semantic and/or morphological similarities. Storjohann provides a perspective framed by considering how lexical items may be stored in the mental lexicon; her investigation begins with a look at the lexicographic treatment of paronyms over the past years in German dictionaries, challenging critiques of past and present dictionaries treating easily confused German words, produced mostly for the benefit of the uneducated and the foreign language learner. She finds in some cases a prescriptivism apparent in the choice of illustrative quotations that is revealed by the dependence on a narrow range of "literary" sources. Yet, as Storjohann makes clear, "[t]oday, rather than consulting dictionaries, or possibly because users encounter dictionary problems, people turn to online forums and ask the language community for advice".

It is important to note that even as the research is initially based on the digital faculties available for sorting and compiling, organizing the work will depend on the natural faculties of the human operator, that is, the analysis will not be automated but qualitatively described and explained by a research team. The challenge for lexicographers will be to "make dictionaries more appealing to users" by broadening lexicographic possibilities through historical contemplation of dictionary artifacts leading to "imaginative approaches."

Historical research has led her to conclude that "new lexical approaches" are necessary and that there is also a need for the production of empirically based reference guides to present-day German paronyms. And, as new perspectives are formed by a critical study of prior studies and dictionaries, a working definition of ‘paronym’ must be developed to provide a foundation for the historical and also contemporary research.

Io Manolessou and Georgia Katsouda provide a contrasting perspective on historical lexicography with their article that describes the long-term project publishing The *Historical Dictionary of Modern Greek, both of the Standard Language and the Dialects* starting in 1933. In many ways the project reflects work on other historical dictionaries in the familiar nationalist and romantic ideologies of the nineteenth century; naturally, a historical dictionary of Greek from that perspective would be fulfilled by showing a direct linguistic and cultural lineage with Ancient Greek civilization. However, simplifying matters somewhat, Greek developed an increasingly divergent linguistic history from the first centuries AD which requires attention to the dialectal developments as sources for the internal reconstruction of Modern Greek. Therefore, the standard would be complemented with a record of all its dialects. As Manolessou and Katsouda point out, “it was and remains a necessary decision [including the dialects], rendered inevitable by the special sociolinguistic circumstances of Greek linguistic history.”

Manolessou and Katsouda stress the importance of the project’s use of primary sources rather than secondary lexicographic ones. This marks the contrasting perspective of the lexicographic account of Modern Greek being published by the Academy of Athens in that other Modern Greek dictionaries avoid “literary” language because it is considered “deviant” or “artificial”. Of course, the task at hand is magnified by the collection and sorting of primary sources that record a multiplicity of dialects. As the two authors emphasize, citing the words of Henry and Renee Kahane, part of the benefit of this method has the consequence of presenting “the wealth of forms and meanings, of derivations and metaphorizations which a given lexeme underwent in the many dialects, [which is] turned into an impressive display of linguistic variation and creativity.” Of course, all that linguistic variation and “exhibition of the genius of the Greek language” creates challenges and opportunities not generally encountered in other historical lexicographic projects. Manolessou and Katsouda do, however, call our attention to the unfortunate similarity to other ambitious lexicographic endeavors of our time in the shortfall of required funding and personnel to undertake a painstaking and time-consuming adventure in the linguistically based excavation and presentation of cultural heritage.

Kateřina Voleková in her article, “The Importance of the Bible Translation for the Old Czech Lexicography,” provides a detailed biobibliographical account and critical commentary on how the history of translations of the Bible and related liturgical texts form an indispensable record for the historical lexicography and lexicology of the Czech language. In support of her description of the methods and procedures established for the serial production of the *Electronic Dictionary of Old Czech*, (*Elektronický slovník staré češtiny* 2006), she reminds us of the essential relationship between critical editions of literary artifacts and an evidence-based lexicography, and the unique challenges of textual and lexicographical scholarship attendant to particular linguistic and literary communities. In this case, collaborative editorial work on Old Czech and Early Modern Czech within the Czech Language Institute supports the work of their colleague lexicographers. Because there is a rich tradition associated with Czech translation of the Bible, the first one being completed in the middle of the fourteenth century, it is possible to trace linguistic change over the centuries.

The history of the translations shows attempts to find Czech equivalents for expressions in the Latin editions of the Bible. In this way, morphological, phonological and semantic evidence can reveal a shift from older, even obsolete forms of the language to the contemporaneous language of the time, but with the proviso that the equivalents may represent the influence of the source language on word formation rather than principles of Czech morphology. Because there is a multiplicity of translations, of the whole Bible, books of the Bible, psalters, and the like, and that much of this output is only available in manuscript or photocopy, there is a commitment to digitizing the data, piece by piece, for the greater ease of identifying, collecting and analyzing the linguistic evidence. Voleková describes the variety of studies—digital, textual and linguistic studies—of the history of Czech editions of the Bible that have been undertaken previously and that are part of the ongoing painstaking effort of producing digital critical editions, both of the texts themselves, the history of the *Dictionary of Old Czech*, and for the resultant and continuing work on the *Electronic Dictionary of Old Czech*.

The next article provides another perspective on collaborative research and text production by scholars from the Department of Language Development of the Czech Language Institute of the Czech Academy of Sciences. **Irena Fuková** and **Markéta Pytlíková** offer a critical overview of the developmental stages of the *Elektronický slovník staré češtiny* (*Electronic Dictionary of Old Czech*). Fuková and Pytlíková provide pertinent and supporting evidence drawn from textual and linguistic analyses of the painstaking work collecting and recording data on index cards performed by the earlier research team of Jan Gebauer that underwrote

Slovník staročeský (*Dictionary of Old Czech*), 1903-1916. They add a critical description of the lexicographic depth of *Staročeský slovník* (*Old Czech Dictionary*), 1968-2008, and discuss the limited usefulness of *Malý staročeský slovník [Little Old Czech Dictionary]*,1978, which covers parts of the vocabulary untreated by the other two.

In this article we see the alignment and reconfiguration of principles and lexicographic guidelines across more than a century of scholarship and dictionary production. Partly this is necessary because the history of historical dictionaries of Old Czech is shown to be a combination of unfinished lexicographic work of rich descriptive quality along with dictionaries that can either sometimes provide unreliable entries or in another case relatively minimal lexicographic documentation.

The current effort balances the need to produce a completed record of Old Czech that will be useful for scholars and will also be accomplished in a timely fashion. The research and lexicographical work on the *Electronic Dictionary of Old Czech* combined with the digitized *Old Czech Dictionary* has the object of providing a complete lexicographically sound historical accounting and description of the vocabulary of Old Czech from the twelfth to sixteenth centuries.

Ana Mihaljević in her contribution, "Latin and Italian Influence on Croatian Dictionaries up to Jakov Mikalja's *Thesaurus Linguae Illyricae"* fore-grounds what for some may seem a controversial point, or at least an opportunity for a reconsideration of conventional knowledge regarding what types of dictionaries occurred first. It is not so much what is a first, or second dictionary type, but more a case of what are the cultural and societal pressures that call forth a dictionary of any type.

And since we have an incomplete record of manuscript and even print artifacts that we would consider as part of a history of dictionaries, making pronouncements about firsts of any extant text is always fraught with speculative provisos. Is it the monolingual tradition in lexicography or the bilingual tradition in lexicography (including multilingual lexicography)? A global perspective, as yet not fully realized or familiar, suggests further questioning of the more localized historical certainties, or at least conventions, about the history of lexicography.

Another interesting point in this history of Croatian lexicography is the influence of non-first language speakers on the recording of a variety of Croatian that has become important in its historical dictionary and word study. And that multi word expressions find their way into a dictionary history is not a later occurrence but an early occurrence, a result of producing multilingual dictionaries for students. We also see another way of looking at the succession of dictionary editions, as additions, and not simply dismissed as a kind of plagiarism; however, we may misunderstand the dissemination of knowledge across time and geography if a change of word order or reformatting an entry is overlooked as a source of linguistic value. Some of 20th century historical research into dictionaries has this tendency to overlook the dictionary as a continuous text; that is each individual dictionary, particular artifact, of a larger possible (but not printed) abstract dictionary. “The dictionary” is abstract and virtual in the sense that it is an unprinted, unwritten dictionary which can only be accessed in discontinuous and particular instances, not in any one printed or written production.

Mihaljević shows us how a nonstandard variety of Croatian can influence the development of a history of ideas as part of pan European literary production. A Lingua Franca becomes the vehicle of vernacular diaspora. And we also note the early occurrence of lexicography in a non-romantic linguistic community, a case of language contact through commerce and the influence of administrative and intellectual registers of Greek and Latin. It is important to be reminded always that we really cannot say with any certainty that there was not already a lexicographic tradition in manuscript form that would record one or another variety of Croatian before the records of which we are now aware.

The next article also explores the variety of societal and political dimensions of finding local translation equivalents, in this case, Tajik for Russian. The lexicographic decisions of producing technical vocabularies for a linguistic community are described and reflected upon by **Abdusalom Mamadnazarov** in his “A Select History of Modern Russian-Tajik Special Lexicography.” His study focuses on bilingual Russian-Tajik terminological dictionaries, also calling attention to the much more complex and rich tradition of tracing the borrowing of terms through the influence of Russian technical register from other historically developed technological vocabularies such as French, English and German. Currently, with a gradual shift of the lingua franca from Russian to English the history of Tajik lexicography holds great interest for studies of word-formation and translation equivalents. Mamadnazarov explains that “[a]ccording to the majority of Tajik linguists, in the first years of the Soviet period the words and expressions of the Russian language were borrowed, very often, without any effort to find their translation or equivalents.”

It is clear that the standardization of terminology in the Tajik language requires both a synchronic and diachronic approach to producing dictionaries for comprehension and for writing which includes not only following current texts that may abandon Russian influenced borrowings for a less widely understood Arabic influenced expression. Mamadnazarov calls attention to the variety of explanatory and descriptive specialized dictionaries produced through the successive eras of the twentieth and twenty-first centuries with explications of the strengths and shortcomings of the micro- and macrostructures of some of the more important contributions to Russian-Tajik lexicography of technical terms. He asserts the importance of these standardizing technical dictionaries in the wider effort to promote an impetus to augment the production of a more comprehensive Tajik bilingual tradition of dictionary production and use.

Frits van der Kuip, Janneke Spoelstra and Willem Visser present a detailed historical overview of the development, successes and trials of the lexicography of Frisian. One area, historical lexicography, remains relatively underdeveloped and thus is a research obstacle to not only the history of Frisian but also as a basis for considering the linguistic description of present and past varieties of Frisian. As the authors say, "…interest in Frisian studies appears to have shifted to synchronic linguistic research…" Even so, Frisian has a richly developed lexicographic tradition. As a minority language in the Netherlands, it is not surprising that bilingual lexicography of Dutch-Frisian has been a main occupation of dictionary makers. Perhaps, the influence of the current work on a digital Frisian dictionary will aid in raising the level of consciousness among funding agencies for more historical work.

Though for many years Frisian was mostly a spoken language, as we have already seen in the paper by Deutsch it can be found in legal documents up to the end of the Middle Ages. Furthermore, the attempt at establishing a standardizing dictionary as noted in articles included in this collection is not uncommon for minority languages and for dialects, or the common vernacular, to have not supplied enough of a historical record to support the production of a standard or standard influencing dictionary of the languages. With that effort comes questions of whether the lexicographical project may be helping to create a new standard; trying to revive a language; elevating a variety of a language into a standard; obstructing linguistic change; or emphasizing description of an existing standard (see Zgusta 2006, 2 and 186-95).

This state of affairs of the lesser support for historical lexicography, duly noted in a variety of chapters in this volume—not confined to minority languages, but also at work in the editorial rooms of comprehensive national dictionaries of world languages, the OED for example—tells us that at the moment scholarship on the history of dictionaries and word studies is in a time of broadening perspectives and narrowing possibilities.

3. A List and Associated Bibliography of Previous ICHLL Meetings, 2002-2021

ICHLL 1 – Leicester, England (2002)
ICHLL 2 – Gargnano del Garda, Italy (2004)
ICHLL 3 – Leiden, The Netherlands (2006)
ICHLL 4 – Edmonton, Canada (2008)
ICHLL 5 – Oxford, England (2010)
ICHLL 6 – Jena, Germany (2012)
ICHLL 7 – Gran Canaria, Spain (2014)
ICHLL 8 – Bloomington, Indiana, USA (2016)
ICHLL 9 – Santa Margherita Ligure, Italy (2018)
ICHLL 10 – Fryske Akademy, Leeuwarden-Ljouwert, Netherlands (2019)
ICHLL 11 – University of La Rioja, Logroño, Spain (2021)
ICHLL 12 – The Université de Bretagne Sud, Lorient, France (2022)

Julie Coleman and Anne McDermott, eds., *Historical Dictionaries and Historical Dictionary Research* (De Gruyter, 2002)

John Considine and Giovanni Iamartino, eds., *Words and Dictionaries from the British Isles in Historical Perspective* (Cambridge Scholars Publishing, 2007)

Marijke Mooijaart and Marijke van der Wal, eds., *Yesterday's Words: Contemporary, Current and Future Lexicography* (Cambridge Scholars Publishing, 2008)

John Considine, ed., *Adventuring in Dictionaries: New Studies in the History of Lexicography* (Cambridge Scholars Publishing, 2010)

John Considine, ed., *Current Projects in Historical Lexicography* (Cambridge Scholars Publishing, 2010)

John Considine, ed., *Webs of Words: New Studies in Historical Lexicology* (Cambridge Scholars Publishing, 2010)

Charlotte Brewer, ed., *The Fifth International Conference on Historical Lexicography and Lexicology* (Oxford University Research Archive, 2010)

Bettina Bock and Maria Kozianka, eds., *Weiland Wörter-Welten – Akten der 6. Internationalen Konferenz zur Historischen Lexikographie und*

Lexikologie (Jena, 25. – 27. Juli 2012). Whilom Worlds of Words – Proceedings of the 6th International Conference on Historical Lexicography and Lexicology (Verlag Dr Kovac, 2014)

Victoria Domínguez Rodríguez, Alicia Rodríguez Álvarez, Gregorio Rodríguez Herrera, and Veronica C. Trujillo Gonzaléz, eds., *Words across History: Advances in Historical Lexicography and Lexicology*, (Servicio de Publicaciones y Difusión Científica de la Universidad de Las Palmas de Gran Canaria, 2016)

Michael Adams, ed., "Problems of Chronology in Historical Lexicography and Lexicology", (*Dictionaries*, 39:1, 2018)

Chiara Benati and Claudia Händl, eds., *From Glosses to Dictionaries: The Beginnings of Lexicography* (Cambridge Scholars Publishing, 2019)

Chapter One

Lexicographical Planning and Official Ideology: The Case of China's National Plans for the Compilation and Publication of Lexicographical Works

Chiara Bertulessi

Abstract

In the history of the People's Republic of China, the country's political leadership has always attached great importance to developing its lexicographical sector and to the role of dictionaries within society. Since the mid-1970s, the Chinese government has issued and implemented three National Plans in the field of lexicography (1975-1985; 1988-2000; 2013-2025), intended to promote the compilation, revision, and publication of reference works in the country.

Adopting the theoretical framework of critical lexicography, the present article addresses the topic of the interplay between lexicographical activities and ideology. Specifically, it proposes an analysis of three official documents that were issued by relevant Chinese organizations to introduce and announce the implementation of each lexicographical Plan. The key purpose of the article is thus to discuss, through the analysis of the selected documents, the ways in which the official ideology and the leadership's goals participate in framing the guiding principles and objectives of lexicographical activities in the People's Republic of China.

1. Introduction: Theoretical framework and main objectives

Chinese language lexicography can draw on an ancient and well-established tradition. Studies in the field show that proto-lexicographical works existed in China in pre-imperial times, mainly in the form of primers compiled for the purpose of learning characters. Works dating to the Han Dynasty (206 BC-220 AD), like the *Erya* 尔雅 (third century BC) and the *Shuowen jiezi* 说文解字 (second century AD), are generally acknowledged as the oldest-known dictionaries of the Chinese language establishing the principles of Chinese lexicography (Xue 1982; Creamer 1986, 1991; Ming 2006; Yong and Peng 2008). From the Han Dynasty to the twentieth century, lexicography blossomed in China, with the compilation, among many others, of monumental works like the *Kangxi zidian* 康熙字典, the eighteenth-century dictionary commissioned by Emperor Kangxi of the Qing Dynasty (1644-1911) (Yong and Peng 2008).

After the founding of the People's Republic of China (1949) (hereinafter PRC), the new political leadership guided by the Chinese Communist Party engaged in a national project for language reform. Also relying on the efforts already put forth by Chinese intellectuals in the previous decades, in the second half of the 1950s the leadership defined the three main objectives of such reform: the standardization and popularization of the 'common language' or *putonghua* 普通话, the simplification of the Chinese characters to boost literacy, and the creation of a new romanization system for the Chinese language (*pinyin* 拼音) (De Francis 1950; Yin and Baldauf 1990, 282). With language standardization, lexicographical activities acquired even greater importance in the country: language dictionaries, in particular, took on a central role in the definition and prescription of the linguistic standards of the common language (Guowuyuan 1956).

The present article addresses the topic of lexicographical planning in the PRC. Specifically, the article examines the three National Plans issued since the mid-1970s by the Chinese government to boost and regulate the compilation and publication of lexicographical works in the country. Through the analysis of three documents published by the relevant governmental agencies to introduce the publication of the Plans, I focus on the interplay between lexicography and ideology. Foregrounding the role of the official ideology within a specific political and historical context, the theoretical framework of critical lexicography is here adopted. By critical lexicography, I refer to the research field that falls within the scope of metalexicography and which aims to study and assess the function that lexicographical activities and works play in the reception, consolidation, and reproduction of the dominant ideology (Fairclough 1989) and, generally speaking, of a specific representation of reality within a given context (Benson 2001; Hornscheidt 2008; Rodríguez Barcia 2012; Chen 2015).[1] As Joshua Fishman (1995, 34) puts it, "we must interpret dictionaries in context and see them both as resultant of and constructive of their contexts".

Following Hodge and Kress (1979, 6), ideology is defined as "a systematic body of ideas, organized from a particular point of view", a concept that in the two scholars' view also includes political ideologies. Moreover, relying on Reisigl and Wodak's (2009, 88) definition of the concept, ideology—and, in this case, official ideology—is understood "as an (often) one-sided perspective or world view composed of related mental representations, convictions, opinions, attitudes, and evaluations, which is shared by members of a specific social group". Thus, in this article I define official ideology as the "systematic body of ideas" intended as contributing to the maintenance and the (discursive) legitimization of the leadership's political power in the PRC, organized from the point of view and shared by members of the Chinese political leadership.

Therefore, the main purpose of the article is to discuss the ways in which the official political ideology participates in framing both the guiding principles and the objectives of lexicographical activities in the PRC. In the paragraphs that follow, I will show how dictionaries, and reference works in general, are always presented in the documents analyzed both as tools in service of and influenced by not only cultural, linguistic, and social needs, but also ideological and political needs. Given the different historical moments in which the Plans were issued (1975, 1988 and 2013), the analysis also adopts a diachronic perspective to identify the changes in the ways in which the political objectives associated to the field of lexicography have changed since the mid-1970s. In this regard, I argue that the analysis of the selected sources can contribute to the discussion on the functions that the political leadership of the PRC has attributed to the compilation and the publication of reference works both in the past and in the present time.

As will be shown below, despite the different historical periods in which they were published, the language employed in the documents introducing the National Plans can, in many respects, be considered a manifestation of what Michael Schoenhals (1992) has defined as "formalised language", considered as a "form of power managed and maintained by the state". This refers to the consistent use of a "restricted code" by Party

---

[1] Critical lexicography should here be understood as related, but different in purpose and approach, to lexicographical or dictionary criticism, the latter being mainly concerned with the qualitative analysis of lexicographical works (Hornscheidt 2008; Rodríguez Barcia 2012; Nielsen 2018).

officials and in Chinese official political discourse in general. A distinctive feature of Chinese formalized language, which was largely shaped during the Maoist period, is the recurrent resort to “formulations”, or *tifa* 提法, fixed formulations and slogans that represent the correct and prescribed “ways of putting things”, or terminology, to be employed in political discourse (Schoenhals 1992, 1 ff.; Link 2008, 274-278).

Although it is beyond the scope of this article to discuss the meanings and the ideological implications of the official formulations included in the documents, their presence in the selected resources will be analyzed in relation to the function that these assume in framing, in the Plans, the role of lexicographical activities in the country. Specifically, these official formulations will be regarded as the manifestations, on the discursive and linguistic level, of the official ideology (as defined above in this paragraph) in the different historical and socio-political contexts in which the Plans were drafted and issued.

2. China’s national lexicographical plans: An overview

Since the founding of the PRC and, specifically, since the mid-1970s, the Chinese government has promoted three National Plans in the field of lexicography. The purpose of the Plans has been to boost activities aimed at the compilation and the publication of lexicographical works in the country and they have been issued by governmental agencies in charge of the administration and regulation of the country’s publication and media activities.

2.1 The First (1975-1985) and Second Plan (1988-2000)

The first Plan was issued in 1975, following a significant symposium held in Guangzhou between May 23 and June 17 of the same year. The event was convened by the State Publishing Administration (*Guojia chuban shiye guanliju* 国家出版事业管理局, “SPA”) and the Ministry of Education (*Jiaoyubu* 教育部), with the approval of the State Council (*Guowuyuan* 国务院) and was given the title of “Symposium for the Planning of Compilation and Publication of Chinese and Foreign Language Dictionaries” (*Zhongwai yuwen cidian bianxie chuban guihua zuotanhui* 中外语文词典编写出版规划座谈会)[2] (SPA 1975, 251). The symposium gathered more than a

[2] In modern Chinese there is not a unique term that corresponds to the English words “lexicography” and “lexicographical works”. Based on the common usage of

hundred people, including members of China's cultural and educational circles, publishing houses, scholars, and experts in the relevant fields, as well as ordinary citizens, or, as stated in the document, representatives of "workers, peasants and soldiers" (*gong nong bing* 工农兵) (SPA 1975, 251). The first Plan stipulated the compilation and publication of 160 language dictionaries, including both monolingual Chinese dictionaries and bilingual and foreign language dictionaries, to be published between 1975 and 1985. This ten-year Plan was expected to be carried out in two phases: the first (1975-1980) was intended to lead to the publication of 25 Chinese language dictionaries and 78 foreign language dictionaries, while the second (1980-1985) was aimed at the publication of 6 Chinese language dictionaries and 50 foreign language dictionaries (SPA 1975, 253). Among the 160 dictionaries to be published, new dictionaries outnumbered those which were only to be subject to revision. Moreover, the number of foreign language dictionaries expected to be published during the implementation of the first Plan (1975-1985) was considerably higher than the number of Chinese language dictionaries (SPA 1975, 256-265; Wei 2015, 2-3).[3] Wei Xiangqing (2015, 2-3), who has conducted extensive research on the Plans and the development of contemporary Chinese lexicography (Wei et al. 2014; Wei 2015), argues that the leadership's choice to plan more foreign language dictionaries than Chinese language works should be assessed in relation to the political and historical situation of China during that period. China's political instability had a negative influence on lexicographical activities and the publishing sector, and the number of foreign language dictionaries was extremely limited at the time. The decisions adopted during the 1975 symposium were thus aimed at reducing the gap in the different types of dictionaries available to Chinese users. It should also be noted that the 1975 symposium was held at a time of strong social and political instability, also due to the negative impact that national political campaigns like the Cultural Revolution and the "Criticize Lin, Criticize Confucius"

lexicographical terminology by contemporary Chinese scholars, in the present article I use the word "dictionary" to translate the Chinese *cidian* 词典 (or "word dictionary", which is mainly to refer to language dictionaries) and "lexicographical works" to translate the more general term *cishu* 辞书.

[3] Among these, the main foreign languages involved were English, Russian, French, German, Japanese, Spanish, along with North Korean language, Arabic, Vietnamese, Indonesian, Hindi, Greek, and even Esperanto. The languages are listed in decreasing order, based on the number of dictionaries planned for each language and according to the data collected by Wei (2015).

movement (1973-1974)[4] had on the political arena and the cultural and publishing sectors.

The lexicographical sector was, therefore, also affected by the political climate in the PRC. In this context, in the spring of 1974—one year before the Guangzhou symposium was held—the lexicographers engaged in the compilation of the *Xiandai hanyu cidian* 现代汉语词典 (an authoritative prescriptive dictionary of modern standard Chinese, or *putonghua* 普通话)[5] were subject to a flood of harsh criticism promoted by some of the most extremist political figures within the Chinese Communist Party (later known as the Gang of Four), due to the allegedly reactionary contents of the so-called trial edition (*shiyongben* 试用本) of the dictionary, which had been previously distributed in 1973 for review purposes. The criticisms eventually evolved into a denunciation campaign which gave rise to a heated debate among intellectuals and politicians on the tasks and functions of lexicographical activities. As will be shown in more detail in the following paragraphs, although not officially mentioned in the documents, some of the decisions adopted at the 1975 symposium had a direct impact on the activities for the revision and the publication of the first edition of the *Xiandai hanyu cidian* 现代汉语词典 (1978) (Cidian bianjishi 2004, 192; Lee 2014, 431-432; Bertulessi 2019, 60-70 and 2020).

---

[4] The so-called "Criticize Lin, Criticize Confucius" (*Pi Lin pi Kong* 批林批孔) movement was a political campaign launched by Mao Zedong in 1974 and promoted by the so-called 'ultralefitsts' (Jiang Qing, Yao Wenyuan, Wang Hongwen and Zhang Chunqiao, also known as the Gang of Four). The campaign officially targeted former vice-chairman of the Chinese Communist Party and Minister of defence Lin Biao (1906-1971), connecting his ideology to Confucianism, which, according to Mao, represented the backward, reactionary, and oppressive legacy of China's past. However, the campaign was also a less-explicit offensive against prime minister Zhou Enlai (1898-1976). See Goldman (1975), Gregor and Chang (1979).

[5] The *Xiandai hanyu cidian* is a middle-size dictionary of modern standard Chinese compiled by the Institute of Linguistics of the Chinese Academy of Social Sciences. Its compilation was originally planned by the Chinese government as part of the project for language standardization in the second half of the 1950s. During the 1960s and the 1970s, trial editions of the dictionary were printed for a limited circulation to collect reviews and opinions. The events occurred in 1974 and the decisions adopted during the 1975 symposium on lexicography heavily interfered with the publication process. The first official edition of the dictionary was thus not published until 1978. Since then, it has been subject to several revision processes. In 2016, its 7th edition was published by the Commercial Press (*Shangwu yinshuguan* 商务印书馆) and the dictionary continues to be regarded as one of the most authoritative dictionaries of modern standard Chinese (Bertulessi 2019; Lee 2014).

Examples among the many language dictionaries from the first Plan published after 1975 are the revised editions of Chinese monolingual dictionaries as the *Xinhua zidian* 新华字典 (planned to be published in 1977, published in 1979) or the *Cihai* 辞海 (planned for 1978, published in 1979), or newly compiled bilingual dictionaries as the *Xin Ying Han cidian* 新英汉词典 (1975, English-Chinese) or the *Han E cidian* 汉俄词典 (planned for 1976, published in 1977, Chinese-Russian).[6]

Thirteen years later, in November 1988, a second national symposium on lexicography was convened in Chengdu by the State Administration of Press and Publication (*Guojia xinwen chubanshu* 国家新闻出版署, "SAPP"). The symposium, entitled "Second National Symposium for the Planning of Compilation and Publication of Lexicographical Works" *(Di er ci quangguo cishu bianxie chuban guihua zuotanhui* 第二次全国辞书编写出版规划座谈会), led to the compilation of a new Plan, which included a list of 169 lexicographical works, to be published between 1988 and 2000. Unlike the first (1975-1985), which focused solely on language dictionaries, the new Plan was also intended to increase the number of other types of lexicographical works, such as specialized dictionaries and encyclopaedias, both in the fields of social and natural sciences (SAPP 1989; Chi 1990; Wei et al 2014, 27-28; Wei 2015, 4). The 1988-2000 Plan was also designed to be implemented in two phases (1988-1993 and 1994-2000). The document introducing this Plan stresses that a new strategy was needed to continue the work started between 1975 and 1985 and to boost the compilation of lexicographical works, which should also adapt to the "new situation of China's reform and opening up" (*Wo Guo gaige kaifang de xin xingshi* 我国改革开放的新形势) (SAPP 1989, 473).[7] The 1988 Chengdu symposium on lexicography and the Plan that followed should thus be regarded as an integral part of the reform and modernization era, which also involved the

---

[6] For the complete list of the titles included in the first Plan see SPA 1975 [2013], 256-265.

[7] In 1978, the political leadership of the post-Maoist period—*de facto* guided by Deng Xiaoping—had officially embarked on a path of economic reforms, also known as "reform and opening up". The Third Plenum of the Eleventh Congress of the Chinese Communist Party, held in December 1978, is generally acknowledged as the political event that sanctioned the upholding of economic reforms and a more general shift from the politics of the era of Mao Zedong, who had died in September 1976.

development of China's social and natural sciences, focusing on training, education, and the nurturing of talents (Deng 1977 and 1978; Cogan 1980).[8]

Examples of dictionaries planned and published between 1988 and 2000 are Chinese monolingual dictionaries like volumes of the monumental *Hanyu da zidian* 汉语大字典 (1989-1990), the *Xin ci xin yu cidian* 新词新语词典[9] (1989), bilingual dictionaries as the *Xin Han Ri cidian* 新汉日词典 (1991, Chinese-Japanese) or specialized dictionaries as the *Zhongguo wenxue dacidian* 中国文学大辞典[10] (planned to be published in 1995, published in 1997).[11]

According to Wei Xiangqing (2015, 1-2), the first two Plans succeeded in promoting the development of lexicography in the PRC, also contributing to the revitalization of the publishing sector in post-Maoist China. In particular, the first Plan laid the foundations for the development of lexicographical activities aimed at the compilation of both Chinese and non-Chinese language dictionaries. While the second Plan (1988-2000) was also targeted at the publication of language dictionaries, its main function was to expand the activities in the field of specialized, non-language-oriented lexicography, promoting the publication of other types of reference works.

2.2 The Third Plan (2013-2025)

The third and latest Plan was issued by the State Administration of Press, Publication, Radio, Film and Television (*Guojia xinwen chuban guangdian zongju* 国家新闻出版广电总局, "SAPPRFT")[12] in 2013, and is expected to conclude in 2025. Since 2013, the Plan has been subject to two revisions, respectively in 2016 and 2017.

The 2013 document officially launching the Plan outlines the publication of 189 lexicographical works, belonging to different typologies and including both revisions of existing works (151) and new compilations (38) (SAPPRFT 2013). A peculiarity of the new Plan is that it also provides for works to be published in a digital version, along with the printed

[8] Deng Xiaoping repeatedly insisted on the valorization of education and the nurturing of talents as being crucial to the success of China's modernization (Deng 1977, 1978).

[9] A dictionary of neologisms.

[10] A dictionary of Chinese literature.

[11] For the complete list of the titles included in the second Plan see SAPP 1989, 475-478.

[12] In 2018 SAPPRFT was dismantled and replaced by the National Radio and Television Administration (NRTA).

reference works. Although representing only a small part of the 189 reference works listed in the original 2013 Plan (8 out of 189), the decision to include digital works attests to the growing interest of the Chinese lexicographical and publishing sector towards digitization,[13] an interest that is also stated in the document introducing the Plan (SAPPRFT 2013).

It should be noted that, following the 2016 and 2017 revisions, the third Plan now includes a total of 235 works to be published by 2025 (SAPPRFT 2016, 2017). According to official data, 30% of the planned lexicographical projects had already been completed by 2017 (SAPPRFT 2017). As of 2021, examples of already published reference works include revised editions of authoritative Chinese monolingual language dictionaries as the *Xiandai hanyu cidian* 现代汉语词典 (7th edition, 2016) and *Cihai* 辞海 (7th edition, 2020), along the with newly compiled bilingual language dictionaries and specialized monolingual or bilingual dictionaries and encyclopaedias, such as the *Yuyanxue yu yingyong yuyanxue baike quanshu* 语言学与应用语言学百科全书[14] (planned for 2016, published in 2017).[15]

Finally, in discussing the topic of lexicographical planning in the PRC, the role of the Chinese Association for Lexicography and of publishing houses in determining the reference works planned for compilation and publication should also be mentioned, especially as regards to the third and latest Plan. The Chinese Association for Lexicography (*Zhongguo cishu xuehui* 中国辞书学会, or ChinaLex) is a national association founded in 1992.[16] Although the involvement of ChinaLex is not explicitly mentioned in the documents introducing the 2013 lexicographical Plan and its 2016 and 2017 revisions, the same documents and other sources do mention the contribution given by experts in the process of examination and assessment of the lexicographical editorial projects proposed by the publishing houses to be included in the 2013-2025 (SAPPRFT 2013, 2016, 2017; Jiang 2014, 6). Moreover, the charter of Chinalex (1999) also lists the “formulation of lexicographical planning and provision of consultation to the relevant

---

[13] See Wei et al. (2014, 177 ff.) on the development of Chinese digital lexicography in the last decades.

[14] An encyclopaedia of linguistics and applied linguistics.

[15] For the complete lists of works included in the third Plan see SAPPFRT 2013, 2016 and 2017.

[16] According to the Association charter (ChinaLex 1999), it is subject to the supervision of China’s Ministry of Education and Ministry of Civil Affairs (*Minzhengbu* 民政部).

national organs" as part of the Association's tasks.[17] The involvement of ChinaLex in the definition of the third Plan and the selection of the dictionaries and references woks to be published between 2013 and 2025 is also mentioned in a report presented by its former president Jiang Lansheng at the tenth annual meeting of the Association held in November 2013 (Jiang 2014). In outlining the future tasks of the Association, Jiang mentions the need to continue to provide support and actively take part in the definition and the implementation of the country's lexicographical planning (Jiang 2014, 6-7). In the report, Jiang briefly describes the process that had led to the publication by SAPPRFT of the new, third Plan, also mentioning the involvement of experts in the selection and evaluation of the lexicographical works to be included in it. Moreover, Jiang (2014, 7) stresses the need for the Association to assist the departments responsible for the plan in its implementation and promotion. To conclude, these elements suggest that members of ChinaLex, as subject experts, are actively engaged in the definition and implementation of the Plan, namely evaluating the applications submitted by publishing houses, which propose their editorial projects to the relevant institutions.

3. Analysis of the selected documents

For the present article, three official resources related to the three National Plans have been selected and analyzed. The first is the report of the 1975 symposium, held in Guangzhou, signed and issued on July 16, 1975 by the State Publishing Administration which, together with the Ministry of Education, was also the organizer of the symposium. The document is entitled "State Publishing Administration Report on the Symposium for the Planning of Compilation and Publication of Chinese and Foreign Language Dictionaries" (*Guoija chuban shiye guanliju guanyu zhongwai yuwen cidian bianxie chuban guihua zuotanhui de baogao* 国家出版事业管理局关于中外语文词典编写出版规划座谈会的报告) (SPA 1975), and is composed of four main sections: an introductory section, a first paragraph titled "Compilation Guidelines and Principles" (*Guanyu bianxie de fangzhen yu yuanze* 关于编写的方针与原则), followed by two other paragraphs, respectively titled "On the Planning of Compilation and Publication Activities and Distribution of Labor" (*Guanyu bianxie chuban guihua he fengong xiezuo* 关于编写出版规划和分工协作) and

[17] The original Chinese wording is "4. 为国家有关部门制定辞书规划和提供咨询" (ChinaLex 1999).

"Strengthening the Leadership of the Party" (*Jiaqiang dang de lingdao* 加强党的领导).

The second selected resource is the report of the November 1988 symposium. Published almost one year later, on July 18, 1989, it is titled "State Administration of Press and Publication Report on the National Plan for the Compilation and the Publication of Lexicographical Works" *(Xinwen chubanshu guanyu quangguo cishu bianxie chuban guihua (1988-2000 nian) de baogao* 新闻出版署关于全国辞书编写出版规划 (1988-2000 年) 的报告) (SAPP 1989). Similarly to the 1975 document, this report is also divided into different sections, the first of which is a non-titled introduction. The following paragraphs are respectively titled "The Rapid Development and Problems of the Compilation and Publication of Lexicographical Works" (*Cishu bianji chuban de xunsu fazhan he cunzai de wenti* 辞书编辑的迅速发展和存在的问题), "Compilation Guidelines and Principles" (*Guanyu bianxie de fangzhen yu yuanze* 关于编写的方针与原则), "Strengthening Planning, Promoting the Compilation and Publication of Lexicographical Works" (*Jiaqiang guihua, tuidong cishu bianxie chuban gongzuo* 加强规划，推动辞书编写出版工作), and "Strengthening the Leadership in the Compilation and Publication Activities of Lexicographical Works" (*Jiaqiang dui cishu bianxie chuban gongzuo de lingdao* 加强对辞书编写出版工作的领导) (SAPP 1989).

Lastly, the third document to be discussed in the present article is not a symposium report, but a notice issued to announce the official publication and enforcement of the third National Plan in the field of lexicography. This notice was published on October 18, 2013 by the State Administration of Press, Publication, Radio, Film and Television (SAPPRFT 2013). Although shorter than the two previous symposium reports, the 2013 notice is also divided into different sections, the first of which is an introduction on the enforcement of the new Plan. Moreover, four paragraphs follow, respectively titled "Scope and Structure of the Plan" (*Guihua de guimo he jiegou* 规划的规模和结构), "Objectives and Features of the Plan" (*Guihua de mubiao he tedian* 规划的目标和特点) and "Requirements for the Implementation of the Plan and Safeguarding Measures" (*Shishi guihua de yaoqiu he baozhang cuoshi* 实施《规划》的要求和保障措施 (SAPPRFT 2013).

Each of the selected documents is followed by the complete list of the lexicographical works planned to be published, along with detailed information of each publication. Specifically, the 1975 and the 1988 Plans include each work's language, title, number of entries, institution(s), or persons in charge of the related lexicographical activities, publisher and expected year of publication (SPA 1975, 256-265; SAPP 1989, 475-478), while the 2013 Plan lists the title, the information about the responsible institutions or persons, the publisher and intended publication form (printed or digital) (SAPPRFT 2013).

Despite their differences, these three documents share several features. First, their structure is similar, presenting an introduction with details about the conditions that led to the compilation of each Plan and their main characteristics. Second, the documents are all divided into more titled paragraphs, treating different aspects of each Plan. Accordingly, a section addressing the topic of the Plan's guiding principles, objectives, and features is included in all three documents, which, as I discuss below, is of particular interest for the objectives of the present article. These documents thus share several similarities in terms of structure and objectives. More importantly, they all contribute to the discursive framing of the tasks that the Chinese political leadership has officially assigned to lexicographical activities in the country in different historical moments. For these reasons, I argue that they can be analyzed and compared and that their analyses should also take into account the more general political and ideological objectives of the historical periods in which they were drafted and (are being) implemented.

By also pinpointing the recurrent inclusion of the official formulations typical of China's political discourse, the aim of the paragraphs that follow is to draw attention to the main ideological elements that emerge within each document, elements which play a role in the definition of both the principles and the objectives that should guide the lexicographical activities provided for in the Plans.

3.1 The First Plan: Language dictionaries and their "class character"

As indicated above, the report introducing the first Plan (SPA 1975) is composed of four sections, the first of which functions as an introduction. The three remaining paragraphs respectively examine and present the guiding principles and the orientation of the compilation of lexicographical works, the planning of compilation and publishing activities and the task division within the project, while the last paragraph is titled "Strengthening the Party's Leadership".

The first (and introductory) section provides information about the Guangzhou symposium and on the intentions and objectives shared by the participants in compiling the first lexicographical Plan (1975-1985). Moreover, it illustrates the reasons and needs underlying the symposium and the writing of the first Plan, also providing an overview of the ideological principles that inspired it. First, the "class character" (*jiejixing* 阶级性) of language dictionaries is presented as a shared belief among the participants to the symposium. The document reads that acknowledging a specific class character to language dictionaries means that different social classes make use of different dictionaries to fit their own political line and to fulfil their own political objectives (SPA 1975, 251). In this regard, it should be observed that the idea of cultural artifacts, and language dictionaries in particular, as having a "class character" became a topic of debate among Chinese linguists and lexicographers at the time, which also involved the more general issue of the class character of language itself (Xie 1979; Min 1979; Bertulessi 2019, 102-105).

Based on these premises, the introduction suggests an opposition between the dictionaries that had been compiled before the Cultural Revolution and the new dictionaries that should be published as part of the implementation of the 1975 lexicographical Plan. The report reads that the lexicographical activities undertaken during the seventeen years (1949-1966) before the launching of the Cultural Revolution had suffered the negative influence of the so-called "revisionist" (*xiuzheng zhuyi* 修正主义) trend in the political arena. Specifically, because of this ideological trend, many of the existing language dictionaries were "permeated by the poison of feudalism, capitalism and revisionism" (*jintou zhe feng, zi, xiu dusu* 浸透着封、资、修毒素) (SPA 1975, 251). Moreover, the report suggests that, despite the efforts made in the years prior to 1975, there was a pressing necessity to create new dictionaries, the contents and features of which should conform to the needs of China's development. In this regard, the compilation of new dictionaries is also explicitly linked to China's modernization. According to the document, the participants of the 1975 symposium agreed on the need for more and better dictionaries to actively contribute to the transformation of China into a "strong modern socialist country" (*shehui zhuyi de xiandaihua qiangguo* 社会主义的现代化强国) (SPA 1975, 252).

The section titled "Compilation guidelines and principles" also acquires a particular significance. By defining the principles and the orientation that lexicographers engaged in the compilation process should adhere to in their work, this section provides substantial information on the ideological standpoint on which the plan was built. Consequently, it clarifies the functions that the political leadership of the time assigned to lexicographical activities and, specifically, to language dictionaries as the products of such activities. First, the key role of the official political ideology in lexicographical activities is emphasized. According to the document, the compilation of a language dictionary should be guided (*zhidao* 指导) by Marxism, Leninism and Mao Zedong Thought. In this respect, it should be noted that the 1975 Constitution of the PRC officially acknowledged Marxism, Leninism and Mao Zedong Thought as the nation's guiding ideologies.[18] Moreover, it is stated that a Marxist standpoint and orientation should be adopted in the process of handling the selection of words to include in the entry list, the compilation of definitions and of usage examples. In the report, the compilation of language dictionaries is thus framed as an activity that should combine both linguistic and political needs. The features of the dictionary itself should be integrated with the need to disseminate (*xuanchuan* 宣传) the official political ideology, as well as with the need to contribute to the critique (*pipan* 批判) of elements that, at the time, were commonly regarded as harmful to the status quo and the objectives of the official party line. Among the elements to be subject to critique, one can see "revisionism", "the bourgeoise" (*zichan jieji* 资产阶级), the "doctrine of Confucius and Mencius" (*Kong Meng zhi dao* 孔孟之道),[19] and the "ideologies of all the exploiting classes" (*yiqie boxue jieji de yishi xingtai* 一切剥削阶级的意识形态) (SPA 1975, 252).

The official ideology can thus be seen as being assigned a double function in the document with respect to lexicographical activities. On one hand, it is expected to guide the lexicographer's choices in the compilation process. On the other, it should be integrated and reflected in the contents of language dictionaries, as the tangible product of such process.

Furthermore, the 1975 report underlines that language dictionaries should serve the masses and, specifically, workers, peasants, and soldiers (*yuwen cidian bixu wei gong nong bing fuwu* 语文词典必须为工农兵服务). At the same time, they should serve those that were officially presented as ideological and political objectives of the Chinese leadership, which are listed in the document in the form of common official formulations: the "socialist revolution" (*shehui zhuyi geming* 社会主义革命), the

---

[18] Article 2 of the 1975 Constitution, adopted on January 17, 1975, states that Marxism, Leninism and Mao Zedong Thought "are the theoretical basis guiding the thinking of our nation" (马克思主义、列宁主义、毛泽东思想是我国指导思想的理论基础) (Quanguo renmin daibiao dahui 1975).

[19] According to Mao Zedong and thus the Chinese Communist Party's official party line of the mid-1970s, the doctrines of Confucius (551-479 BC) and Mencius (372-289 BC) were regarded as a symbol of the legacy of China's decadent past and the old feudal system that the Communist revolution had dismantled. See footnote 4 in this article.

"construction of socialism" (*shehui zhuyi jianshe* 社会主义建设), the "dictatorship of the proletariat" (*wuchan jieji zhuanzheng* 无产阶级专政), as well as the "international struggle against imperialism and revisionism" (*guoji fandi fanxiu douzheng* 国际反帝反修斗争) (SPA 1975).

Another theme that is central in the report is the idea that the Plan should lead to the establishment of a new type of lexicography and, consequently, to the compilation of a new type of language dictionaries. First, the establishment of a new kind of lexicography is presented as involving the relationship with the past and with foreign cultural elements and their influence on Chinese culture and lexicographical activities. Second, it ties together with the relationship between the "scientific character" (*kexuexing* 科学性) and the objectives of lexicographical compilation and the so-called "revolutionary character" (*gemingxing* 革命性) (i.e., political character) of such activity. Third, the new type of lexicography advocated in the 1975 report addresses the role of the masses and experts in the field of lexicography. Concerning the first point, the expression "*gu wei jin yong, yang wei zhong yong* 古为今用，洋为中用" or "make the past serve the present and the foreign things serve China", included in the document, effectively summarizes the official position on the issue. It stresses the belief that cultural and ideological elements connected to China's past and foreign elements should be subject to critique so as to only absorb or make use of those elements that suit the political, lexicographical and linguistic goals of the time. As for the nature of lexicographical activities, the document insists on the need to combine the scientific character of dictionary compilation and its revolutionary character. According to this perspective, dictionaries should aim to convey scientific and linguistic knowledge. However, at the same time, the goals of the revolution should not be forgotten or divorced from such activity. Finally, the report underlines that the masses should be actively involved both in the compilation of dictionaries and their assessment—a statement that is closely related to the wider issue of the role of experts within society (SPA 1975, 252-253).

Although not explicitly mentioned, this last point clearly echoes the 1974 denunciation campaign against the Institute of Linguistics of the Chinese Academy of Social Sciences responsible for the compilation of the *Xiandai hanyu cidian* 现代汉语词典. As mentioned above, within the wider political campaign "Criticize Lin, Criticize Confucius" promoted by the so-called Gang of Four, the compilers of the XHC had been harshly accused of being the mouthpiece of the reactionary political tendencies that aimed at subverting the "new system" and restoring the "old feudal system" that the Maoist revolution had dismantled. This criticism sparked a heated debate on the role of lexicographers (as experts) and lexicography within politics and society and on the nature of dictionaries as linguistic or political tools (Sun et al. 1974; Yu 1974; Han 1978; Lee 2014, 451-452; Bertulessi 2019, 60-70 and 2020). In this sense, the events of 1974 are closely related to the 1975 symposium, which also had a concrete impact on the lexicographical activities for the compilation of the first edition of the XHC. Consequently, a new working group was set up between 1975 and 1976, bringing lexicographers engaged in the compilation of the XHC to work alongside representatives of ordinary citizens and of the army. The group was later dissolved in 1977, following the political changes sparked by the death of the supreme leader Mao Zedong and the consequent arrest of the Gang of Four (Cidian bianjishi 2004, 192; Wei G. 2017).

3.2 The Second Plan: Science, culture, and the role of experts in lexicography

Following the Chengdu symposium convened by the State Administration of Press and Publication at the end of November 1988, a new plan was approved by the State Council and a report on the symposium was published in July 1989.

As mentioned above, this second Plan (1988-2000) was not limited to language dictionaries, but rather provided for the compilation and publication of other kinds of lexicographical works, such as specialized dictionaries and encyclopaedias, in both the field of social and natural sciences. As was the case of the 1975 report, it includes an introductory section which provides an assessment of the results of the 1975-1985 Plan, acknowledging its key role in boosting lexicographical activities in post-Maoist China and discussing some remaining problems in the Chinese lexicographical field and industry (SAPP 1989, 473). Three titled paragraphs follow, focused respectively on the orientation and guiding principles in the compilation and publication of lexicographical works and on the need to strengthen the guidance of lexicographical activities. In the section dedicated to principles, lexicographical works are identified as not only important but also irreplaceable tools in the reception and dissemination of national culture and its developments (SAPP 1989, 474).

Adherence to the leadership's political ideology and the Party's policies is also stressed. Lexicographical works are described as tools that should "serve the people" (*wei renmin fuwu* 为人民服务) and "serve socialism" (*wei shehui zhuyi fuwu* 为社会主义服务). Moreover, explicit reference to important ideological and political concepts of the time is made.

For instance, the document employs formulations such as “to oppose bourgeoisie liberalization” (*fandui zichan jieji ziyouhua* 反对资产阶级自由化) and “construction of the socialist material and spiritual civilization” (*wuzhi wenming jianshe he jingshen wenming jianshe* 物质文明建设和精神文明建设)[20] in relation to lexicographical activities, thus connecting lexicography to some of the key topics of the leadership’s discourse of the time. Furthermore, it is interesting to note that the report also discusses the lexicographical treatment of entries dealing with political issues, by stating that the inclusion of political content in reference works should be subject to careful revision and be handled in strict accordance with official policies, to avoid the publication of works containing political mistakes (SAPP 1989, 474).

Finally, this report is largely focused on the topics of science and culture, to the extent that the construction and dissemination of scientific and cultural knowledge is described as the basic function of reference works. Unlike the 1975 report, a central role is assigned to experts in ensuring the quality of the lexicographical products (i.e., language dictionaries, specialized dictionaries, and encyclopaedias). For instance, the documents states that it is necessary to “rely on experts and scholars” (*yao kao zhuanjia he xuezhe* 要靠专家和学者) and that “efforts must be made to improve the quality of lexicographical works” (*nuli tigao cishu de zhiliang* 努力提高辞书的质量). In this respect, it can be argued that a more general focus on quality, high-academic-level lexicographical activities and the adoption of a scientific approach to lexicographical compilation can be found in the 1989 report (SAPP 1989, 474).

3.3 The Third Plan: Lexicography and cultural soft power

As mentioned above, in October 2013, the State Administration of Press, Publication, Radio, Film and Television of the PRC issued an official Notice on the publication and enforcement of the 2013-2025 national Plan for the Publication and the Compilation of Lexicographical Works.

In its introduction, the document mentions the Eighteenth National Congress of the Chinese Communist Party, held in October 2012, and, specifically, employs the expression “in order to thoroughly implement the spirit of the Eighteenth National Congress of the Chinese Communist Party”

[20] Also known as the “two civilizations” (*liang ge wenxing* 两个文明), these formulations were employed by Hu Yaobang (1915-1989) in his political report to the 1982 Twelfth National Congress of the Chinese Communist Party (Hu 1982; He 2001, 242).

(*wei shenru gunache luoshi dang de shiba da jingshen* 为深入贯彻落实党的十八大精神).[21] This formulation serves as a first framework for the new Plan, along with another wording that was put forward by Hu Jintao, former Party General Secretary and President of the PRC, in his political report delivered at the Eighteenth Party Congress (2012), i.e., "developing a strong socialist culture in China" (*Zhashi tuijin shehuizhuyi wenhua qiangguo jianshe* 扎实推进社会主义文化强国建设)[22] (SAPPRFT 2013).

The focus on (socialist) culture is a recurring theme in the document. In the section dedicated to its main objectives and features, the Plan is framed as a project intended to further boost and improve the quality of lexicographical works and their publishing mechanism and consequently contributing to the country's "cultural soft power" and the development of its "socialist culture". China's "cultural soft power" (*wenhua ruan shili* 文化软实力) is thus mentioned in the notice and officially linked to the lexicographical field and industry.

The Chinese concept of soft power has been described by scholars as a form of recontextualization of the theory of soft power originally developed by American scholar Joseph Nye (1990) as a strategy to be implemented in the field of the foreign policy of the United States (Cao 2014; Riva 2017). Unlike Nye's formulation of the theory, its recontextualization in the Chinese context makes culture the key element in the country's national soft power strategy. As argued by Cao (2014, 176), in relation to China's soft power, culture was "defined not only in terms of its instrumental values in international competition, but more substantially its communicative values in reinforcing national cohesion, identity, and solidarity". The inclusion of *wenhua ruan shili* 文化软实力 in Hu Jintao's political report to the 17th National Congress of the Chinese Communist Party consecrated this official formulation as an integral part of the country's national strategy and the leadership's political discourse (Cao 2014; Riva 2017). Although it is beyond the scope of this article to analyze the complex implications of cultural soft power in relation to the cultural

---

[21] The National Congress, which is held in Beijing every five years, represents the key political event in the life of the Chinese Communist Party. The political report delivered to the Congress by the main leader is regarded by scholars as a primary and significant source to study the official political discourse of the Party, by also analyzing the inclusion and the evolution of specific official formulations (Qian 2012).

[22] This is actually the title of the sixth section of Hu Jintao's 2012 Party Congress Report. See Hu (2012a) for the original Chinese version and Hu (2012b) for the full English translation of the political report.

and publishing industries of the PRC, it should be stressed that the inclusion of this element in the latest Plan shows that lexicography is regarded by the current political leadership (guided by Xi Jinping since 2012) as a component in the enhancement of China's cultural soft power, both domestically and internationally.

The 2013 document stresses the importance of scientific and cultural innovation in different fields and in relation to lexicography, which should also be reflected in the publication of high-quality, specialized, and encyclopaedic reference works (SAPPRFT 2013; Wei 2015, 8). The pursuit of innovation in the lexicographical field is also shown in the call for the digitization of lexicography (*cishu shuzihua* 辞书数字化) and the promotion of projects for the construction of linguistic corpora to be employed in the compilation of reference works (SAPPRFT 2013).

Finally, it should be observed that the idea of lexicography as a field contributing to China's cultural soft power and the focus on innovation and quality are strictly connected to the objective of turning China into a "lexicographical power", a topic which is addressed in more detail in the following paragraph.

4. Towards a "lexicographical power"?

As also argued by Wei (2015), the third lexicographical Plan (2013-2025) is seen as a decisive step in the realization of the strategy that strives to make China a "lexicographical power" or *cishu qiangguo* 辞书强国. This wording was employed by prominent Chinese linguist Li Yuming in 2007, during a speech delivered at Ludong University (Li 2008). In assessing the state of contemporary Chinese lexicography, Li stressed the differences existing between the Chinese lexicographical sector and other well-established lexicographical traditions abroad. In his speech, the linguist identified three criteria that could be adopted in the definition of a "lexicographical power", that is: relying on a strong and competitive international lexicographical market; creating reference works that succeed in acquiring the status of "classics" in the international arena and relying on high quality, academic-level lexicographical theories in the compilation process (Li 2008, 2-3; Wei 2015, 8).

The topic has been the subject of papers written by numerous Chinese scholars and lexicographers, especially since the third Plan was issued in 2013. A recurring theme in the contributions is the evaluation of the weaknesses of the lexicographical sector in the PRC (e.g., Zhang Z. 2010; Zhang Z. and Zhang Q. 2015, 5 ff). Although the undeniable progress and objectives achieved in this sector in recent decades are generally acknowledged (and, in particular, since the launching of economic reforms in 1978), scholars compare it with the characteristics of the countries that are considered lexicographical powers on an international level such as Great Britain or France. Based on the consulted literature, the main steps that Chinese scholars believe that China should undertake in order to transform into a lexicographical power can be summarized as follows: strengthening Chinese lexicographical theory (Li 2008, 3; Zhang Z. 2010; Wang, Zhang H. and Zhang Z. 2014, 91; Zhang D. 2016, 28; Zhang Z. and Zhang Q. 2015, 6-8; Wei 2015, 8); creating academic-level lexicographical works, acknowledged worldwide for their quality, as symbols of the country's "lexicographical power" (Li 2008, 3; Zhang Z. 2010; Su 2012; Wei 2015, 8); valuing talents and relying on internationally authoritative scholars, personalities and groups of experts (Zhang Z. 2010; Wang, Zhang H. and Zhang Z. 2014, 91; Zhang Z. and Zhang Q. 2015, 3-6; Zhang D. 2016, 29) and, finally, ensuring creativity and innovation in the lexicographical field, in contrast to the current predominantly descriptive and repetitive nature of works and research in the field of lexicography (Zhang Z. 2010).

5. Conclusion

The present article has addressed the topic of the planning of lexicographical activities in the PRC since the mid-1970s to the present day. Specifically, it has focused on the ideological elements which emerged in three documents issued by the relevant institutions to officially introduce the Plans, in order to identify the primary guiding principles and the political and ideological objectives that China's leadership has explicitly linked to the compilation, revision, and publication of lexicographical works in the country. The analysis has shown that the lexicographical Plans often reflect the official objectives put forward by China's political leadership, and, more generally, the political climate of the era in which they were compiled. This aspect is also emphasized by the frequent use, in the texts, of official formulations typical of Chinese political discourse.

In this article I did not discuss the issue of the concrete implications and effects that these Plans have had on China's lexicographical field and industry, especially as far as the 1975-1985 and 1988-2000 Plans are concerned. Research on the topic carried out by Wei Xiangqing and her research group at the University of Nanjing shows that Chinese lexicography has seen unprecedented growth since the 1975 Plan was issued and, in particular, since the start of economic reforms at the end of 1978. According to the data collected in Wei et al. (2014) and Wei (2015), between 1978 and 2008, a total of 14,090 lexicographical works was published in the PRC. This figure is 16 times the total of the works that were published between 1949 and 1979, the first three decades since the founding of the country (Wei et al. 2014; Wei 2015, 6).

As for the most recent Plan, the implementation of which will be completed in 2025, I have observed that it is frequently mentioned by Chinese scholars as the one that should turn China into a "lexicographical power". In this respect, it is interesting to note that this interpretation has also appeared in schemes for the periodization of the history of Chinese lexicography, like the one proposed by Zhang Zhiyi and Zhang Qingyun (2015). According to the two scholars, the history of Chinese lexicography can be divided into five phases, of which the third (1950-1977), fourth (1978-2000) and fifth and last (2001-2050?) phase have been respectively identified as the "phase of the small country of lexicography" (*cishu xiaoguo jieduan* 辞书小国阶段), the "phase of the great country of lexicography" (*cishu daguo jieduan* 辞书大国阶段) and, finally, as the phase of the path "towards the lexicographical power" (*zouxiang cishu qiangguo jieduan* 走向辞书辞书强国阶段) (Zhang Z. and Zhang Q. 2015, 1-3).

To conclude, based on the analysis carried out in the present article, it can be argued that lexicography represents a sector of China's cultural industry to which the government and political institutions have attached great importance, both historically and today. The publication of lexicographical works, as an activity planned and regulated by the State, thus proves to be far from neutral. On the contrary, it is frequently framed and presented as an activity to be planned and carried out in service of specific political objectives, the nature of which changed and adapted over decades. In this regard, I argue that the study of the planning and regulation of the activities related to the compilation of lexicographical work within a specific historical, political, and ideological context should be a topic of interest for scholars focusing their research on the field of critical lexicography.

Bibliography

Bertulessi, Chiara. 2020. "'Un minestrone di feudalesimo, capitalismo e revisionismo': lo *Xiandai hanyu cidian* 现代汉语词典 tra rivoluzione, sovversione e reazione". In *Sottosopra. Indagine su processi di sovversione,* Quaderni della Ricerca – 6, edited by Chiara Allocca, Francesca Carbone, Rosa Coppola, Beatrice Occhini., 109-120. Napoli: UniorPress.

Bertulessi, Chiara. 2019. "Analisi critica del discorso lessicografico cinese: uno studio diacronico dello *Xiandai hanyu cidian* 现代汉语词典". PhD diss., Università degli Studi di Milano.

Cao, Qing. 2014. "China's soft power. Formulations, contestations and communication". In *Discourse, Politics and Media in Contemporary China*, edited by Qing Cao, Hailong Tian and Paul Chilton, 171-194. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Chen, Wenge. 2015. "Bilingual lexicography as recontextualization: a case study of illustrative examples in a New English-Chinese Dictionary". *Australian Journal of Linguistics*, 35: 311-333.

Chi, Yan 池言. 1990. "*Guowuyuan pizhun 1998-2000 nian quanguo cishu bianxie chuban guihua*" 国务院批准《1988—2000年全国辞书编写出版规划》 [The State Council approves the1988-2000 national Plan for the Compilation and the Publication of Lexicographical Works]. *Cishu yanjiu* 辞书研究, 1990(4): 1-4.

ChinaLex (*Zhongguo cishu xuehui* 中国辞书学会). 1999. *Zhongguo cishu xuehui zhangcheng* 中国辞书学会章程 [Charter of the Chinese Association for Lexicography]. Accessed January 08, 20201. https://web.archive.org/web/20130824202057/bilex.gdufs.edu.cn/doc /doc-jigou/chinacishu-zhangcheng.htm (archived URL), Accessed January 8, 2021.

Cidian bianjishi 词典编辑室 (Zhongguo shehui kexueyuan yuyan yanjiusuo cidian bianjishi 中国社会科学院语言研究所词典编辑室, Dictionary compilation division, Institute of Linguistics of the Chinese Academy of Social Sciences). 2004. *Xiandai hanyu cidian wu shi nian* 《现代汉语词典》五十年 [Fifty years of the *Xiandai hanyu cidian*], Beijing: Shangwu yinshuguan.

Cogan, John J. 1980. "China's fifth modernization: education". *The Phi Delta Kappan,* 62 (4): 268-272.

Creamer, Thomas B. 1991. "Chinese lexicography". In *Wörterbücher: ein internationales Handbuch zur Lexicographie/ Dictionaries:an International Encyclopaedia of Lexicography, Third Volume*, edited by Franz J. Hausmann, Oskar Reichmann, Herbert E. Wiegand, and Ladislav Zgusta, 2595-2612. Berlin, New York: de Gruyter.

Creamer, Thomas. 1986. "The State of Lexicography in the People's Republic of China: A Brief Survey". *Paper in Linguistics*, 19 (1): 37-53.

De Francis, John. 1950. *Nationalism and Language Reform in China*. Princeton: Princeton University Press.

Deng, Xiaoping 邓小平. 1977. "*Zunshou renshi, zunshou rencai*" 尊重认识，尊重人才 [Respect knowledge, respect talents]. *Renminwang* 人民网, June 10, 2012. Accessed January 08, 2021. https://web.archive.org/web/20120610210156/cpc.people.com.cn/GB/64184/64185/66611/4488675.html (archived URL retrieved on the Internet Archive).

Deng, Xiaoping 邓小平. 1978. "*Deng Xiaoping zai quanguo kexue dahui kaimushi de jianghua*" 邓小平在全国科学大会开幕式上的讲话" [Deng Xiaoping's speech at the National Conference on Education]. *Renminwang* 人民网, January 5, 2006. Accessed January 10, 2020. http://scitech.people.com.cn/GB/25509/56813/57267/57268/4001431.html.

Fairclough, Norman. 1989. *Language and Power*. London and New York: Longman.

Fishman, Joshua A. 1995. "Dictionaries as Culturally Constructed and as Culture-constructing Artifacts: The Reciprocity View as Seen from Yiddish Sources. In *Cultures, ideologies, and the dictionary. Studies in honor of Ladislav Zgusta*, edited by Braj B. Kachru and Henry Kahane, 29-34. Tübingen: Max Niemeyer Verlag.

Goldman, Merle. 1975. "China's Anti-Confucian Campaign, 1973-74." *The China Quarterly*, no. 63: 435-462.

Gregor, A. James, and Maria Hsia Chang. 1979. "Anti-Confucianism: Mao's Last Campaign". *Asian Survey,* 19 (11): 1073-1092.

Guowuyuan 国务院 [State Council of the PRC]. 1956. "*Guowuyuan guanyu tuiguang putonghua de zhishi*" 国务院关于推广普通话的指示 [State Council Directives on the popularisation of *putonghua*]. *Jiangsu jiaoyu* 江苏教育, 1956(04): 3-4 (online version available at: http://www.moe.gov.cn/jyb_xwfb/xw_fbh/moe_2128/moe_2326/moe_1144/tnull_14344.html. Accessed January 8, 2021).

Han, Jingti 韩敬体. 1978. "*Suqing 'liange guji' zai cishu gongzuo zhong de liudu – zhengque pingjia Xiandai hanyu cidian shiyongben*" 肃清 " 两个估计" 在词书工作中的流毒 一 正确评价 《 现代汉语词典》(试用本) [Eliminate the negative influence of "two evaluations" from lexicographical work: a correct review of the trial edition of the *Xiandai hanyu cidian*]. *Renmin jiaoyu* 人民教育, 1978 (11): 10-12.

He, Henry Yuhuai. 2001. *Dictionary of the Political Thought of the People's Republic of China*. London and New York: Routledge.

Hornscheidt, Antje. 2008. "A concrete research agenda for critical lexicographic research within critical discourse studies: an investigation into racism/colonialism in monolingual Danish, German, and Swedish dictionaries”. *Critical Discourse Studies*, 5 (2): 107-132.

Hu, Jintao 胡锦涛. 2012a. “*Hu Jintao zai Zhongguo Gongchandang di shiba ci quanguo daibiao dahui shang de baogao*” 胡锦涛在中国共产党第十八次全国代表大会上的报告 [Hu Jintao’s Report at the 18th Congress of the Chinese Communist Party]. *Xinhuawang* 新华网, November 17, 2012. Accessed January 8, 202, http://www.xinhuanet.com//18cpcnc/2012-11/17/c_113711665_7.htm.

Hu, Jintao. 2012b. “Full text of Hu's report at 18th Party Congress”. *China Daily*, November 18, 2012. Accessed January 8, 2021, https://www.chinadaily.com.cn/china/19thcpcnationalcongress/2012-11/18/content_29578562.htm.

Hu, Yaobang 胡耀邦. 1982. “*Quanmian kaichuang shehui zhuyi xiandaihua jianshe de xin jumian. Hu Yaobang zai Zhongguo Gongchandang di shi er ci quanguo daibiao dahui shang de baogao*” 全面开创社会主义现代化建设的新局面. 胡耀邦在中国共产党第十二次全国代表大会上的报告 [Thoroughly creating new conditions in China's construction of socialist modernization. Hu Yaobang’s report to the twelfth National Congress of the Chinese Communist Party]. Accessed January 8, 2021, https://web.archive.org/web/20190321020057/http://cpc.people.com.cn/GB/64162/64168/64565/65448/4526430.html (archived URL retrieved on the Internet Archive).

Jiang, Lansheng 江蓝生. 2014. “*Wei shixian cishu qiangguo meng er kaituo qianqin – Zhongguo cishu xuehui di si jie ligonghui gongzuo baogao*” 为实现辞书强国梦而开拓前进——中国辞书学会第四届理事会工作报告 [Forging ahead to fulfill the dream of lexicographical power—Report of the 4th council of ChinaLex]. *Cishu yanjiu* 辞书研究, 2014 (1): 4-7.

Kress, Gunther, and Robert Hodge. 1979. *Language as Ideology*. London, Boston, Henley: Routledge & Kegan Paul.

Lee, Siu-Yau. 2014. “Defining Correctness: The Tale of the ‘Contemporary Chinese Dictionary.’” *Modern China*, 40 (4): 426-450

Li, Yuming 李宇明. 2008. “*Nuli fazhan wo guo de cishu shiye – zai hanyu cishu zhongxin jiepai yishi shang de jianghua*” 努力发展我国的辞书事业——在汉语辞书研究中心揭牌仪式上的讲话 [Striving to develop Chinese lexicographical activities. Speech at the inaugural ceremony of the Research Center on Chinese Lexicography]. *Ludong* *daxue xuebao*: *zhexue shehui kexueban* 鲁东大学学报: 哲学社会科学版, 25 (2): 1-3.

Link, Perry. 2008. *An Anatomy of Chinese. Rhythm, Metaphor, Politics*. Cambrige: Harvard University Press.

Min, Jiaji 闵家骥. 1979. "*Tantan yuwen cidian de jiejixing wenti*" 谈谈语文词典的阶级性问题 [On the class character of language dictionaries]. *Cishu yanjiu* 辞书研究, 1979 (1): 48-59.

Ming, Li. 2006. "Chinese lexicography". In *Encyclopedia of Language & Linguistics*, edited by Keith Brown, 362-365. Elsevier.

Nielsen, Sandro. 2018. "Dictionary criticism." In *The Routledge Handbook of Lexicography*. Routledge, edited by Pedro A. Fuertes-Olivera, 78-90. London, New York: Routledge.

Nye, Joseph. 1990. *Bound to Lead: The Changing Nature of American Power*. New York: Basic Books.Qian, Gang. 2012. "Watchwords: The Life of the Party". *China Media Project*, September 10, 2012. Accessed January 8, 2021,
http://chinamediaproject.org/2012/09/10/watchwords-the-life-of-the-party/.

Quanguo renmin daibiao dahui 全国人民代表大会 [National People's Congress]. 1975. Zhongguo renmin gongheguo xianfa 中华人民共和国宪法 [Constitution of the People's Republic of China]. Accessed January 8, 2021.
http://www.law-lib.com/law/law_view.asp?id=343216.

Reisigl, Martin, and Ruth Wodak. 2009. "The Discourse-Historical Approach". In *Methods of Critical Discourse Analysis*, 2nd edition, edited by Ruth Wodak and Michael Meyer, 87-121. Los Angeles: SAGE.

Riva, Natalia. 2017. "Soft Power Narrative in Contemporary China". In *Media and Politics. Discourses, Cultures and Practices*, edited by Bettina Mottura, Letizia Osti and Giorgia Riboni, 241-263. Newcastle upon Tyne: Cambridge Scholars Publishing.

Rodríguez Barcia, Susana. 2012. "El análisis ideológico del discurso lexicográfico: una propuesta metodológica aplicada a diccionarios monolingües del español". *Verba*, 39: 135-159.

SAPP (State Administration of Press and Publication, *Xinwen chubanshu* 新闻出版署). 1989 [1991]. "*Guanyu quanguo cishu bianxie chuban guihua 1988-2000 nian de baogao*" 关于全国辞书编写出版规划 (1988 –2000 年) 的报告 [State Administration of Press and Publication Report on the National Plan for the Compilation and the

Publication of Lexicographical Works]. In *Zhongguo chuban nianjian* 中国出版年鉴 1989, 473-478. Beijing: Guojia chuban nianjian.

SAPPRFT (State Administration of Press, Publication, Radio, Film and Television, *Guojia xinwen chuban guangdian zongju* 国家新闻出版广电总局). 2013. "*Guanyu yinfa* 2013-2025 *nian Guojia cishu bainzaun chuban guihua de tongzhi* 关于印发《2013—2025年国家辞书编纂出版规划》的通知" [Notice on the publication of the 2013-2025 national Plan for the Compilation and Publication of Lexicographical Works]. October 30, 2013. Accessed January 8, 2021, https://web.archive.org/web/20150320224120/http://www.gapp.gov.cn/news/1663/159617.shtml (archived URL retrieved on the Internet Archive).

SAPPRFT (State Administration of Press, Publication, Radio, Film and Television, *Guojia xinwen chuban guangdian zongju* 国家新闻出版广电总局). 2016. "*Guojia xinwen chuban guangdian zongju guanyu tiaozheng* 2013-2025 *nian Guojia cishu bianzuan chuban guihua de tongzhi*" 国家新闻出版广电总局关于调整《2013—2025年国家辞书编纂出版规划》的通知 [Notice on the revision of the 2013-2025 National Plan for the Compilation and Publication of Lexicographical Works]. May 20, 2016. Accessed January 8, 2021, http://219.141.187.25/sapprft/contents/6588/297111.shtml.

SAPPRFT (State Administration of Press, Publication, Radio, Film and Television, *Guojia xinwen chuban guangdian zongju* 国家新闻出版广电总局). 2017. "*Guojia xinwen chuban guangdian zongju guanyu gongbu* 2013-2025 *nian Guojia cishu bainzaun chuban guihua de tiaozheng qingkuang de tongzhi*" 国家新闻出版广电总局关于公布《2013—2025年国家辞书编纂出版规划》调整情况的通知" [Notice on the revision of the 2013-2025 national Plan for the Compilation and Publication of Lexicographical Works]. September 7, 2017. Accessed January 8, 2021, http://www.nrta.gov.cn/art/2017/9/7/art_113_34386.html.

Schoenhals, Michael. 1992. *Doing Things with Words in Chinese Politics. Five Studies*. Berkeley: Institute of East Asian Studies, University of California.

SPA (State Publishing Administration, *Guojia chuban shiye guanliu* 国家出版事业管理局 (1975 [2013]) "*Guoija chuban shiye guanliju guanyu zhongwai yuwen cidian bianxie chuban guihua zuotanhui de baogao*" 国家出版事业管理局关于中外语文词典编写出版规划座谈会的报告 [State Publishing Administration Report on the

Symposium for the Planning of Compilation and Publication of Chinese and Foreign Language Dictionaries]. In *Zhonghua renmin gongheguo chuban shiliao* 中华人民共和国出版史料 14 (1966-1976), edited by Zhongguo xinwen chuban yanjiuyuan 中国新闻出版研究院, 251-265. Beijing: Zhongguo shuji chubanshe.

Su, Peicheng 苏培成. 2012. "*Zou xiang cishu qiangguo zhi lu"* 走向辞书强国之路 [The road towards the lexicographical power]. *Cishu yanjiu* 辞书研究, 2012 (5): 76-80.

Sun, Mantun 孙曼屯, Xue Zhigao 薛志高, Zhao Weimin 赵为民, Cheng Mingke 程明科, and Jia Wenhai 贾稳亥. 1974. "*Ping xin ban Xiandai hanyu cidian*" 评新版《现代汉语词典》 [Review to the new edition of the *Xiandai hanyu cidian*]. *Renmin jiaoyu* 人民教育, 1974 (6): 58-59.

Wang, Tiekun 王铁琨. 2007. "*Guifanhua*、*xiandaihua yu cishu qiangguo – Zhongguo cishu shiye fazhan de sikao*" 规范化、现代化与辞书强国——中国辞书事业发展的思考 [Standardisation, modernization, and the lexicographical power – Reflections on the development of Chinese lexicographical activities]. *Cishu yanjiu* 辞书研究, 2007 (1): 2-7.

Wang, Donghai 王东海, Zhang Hui 张晖, Zhang Zhiyi 张志毅. 2014. "*Cishu qiangguo meng Zhengyuan – tan xin cishu guihua de tuijin cuoshi*" 辞书强国梦正圆——谈新辞书规划的推进措施 [The dream of the lexicographical power. On the measures of the new lexicographical plan]. *Zhongguo bianji* 中国编辑, 2014 (5): 90-92.

Wei, Guanyu 魏冠宇. 2017. "*Xiandai hanyu cidian: xilian yi jiazi, zizi jin zhuji*" 《现代汉语词典》：洗练一甲子，字字尽珠玑. *Jinri yuyanxue* 今日语言学, July 09, 2017. Accessed January 8, 2021, http://ling.cass.cn/xszx/xszx_xscg/xszx_xscg_zzysp/201709/t20170906_3630924.html.

Wei, Xiangqing 魏向清 et al. 2014. *Zhongguo cishu fazhan zhuangkuang baogao* 中国辞书发展状况报告: 1978-2008 [Report on the development of Chinese lexicography: 1978-2008]. Beijing: Shangwu yinshuguan.

Wei, Xiangqing 魏向清. 2015. "*Guojia cishu bianzuan chuban guihua de zhanlve dingwei*" 国家辞书编纂出版规划的战略定位 [On the strategic positioning of the national dictionary compilation and publishing Plan]. *Cishu yanjiu* 辞书研究, 2015 (1): 1-9.

Xie, Zili 谢自立. 1979. "*Yuwen cidian you jiejixing ma – yu Jiaji tongzhi shangque*" 语文词典有阶级性吗——与家骥同志商榷 [Do language dictionaries have a class character? A discussion with comrade JiajTi]. *Cishu yanjiu* 辞书研究, 1979 (2): 100-112.

Xue, Shiqi. 1982. "Chinese Lexicography Past and Present". *Dictionaries: Journal of the Dictionary Society of North America*, 4 (1): 151-169.

Yin, Bo, and Richard B. Baldauf Jr. 1990. "Language reform of spoken Chinese". *Journal of Multilingual & Multicultural Development*, 11 (4): 279-289.

Yong, Heming, and Peng Jing. 2008. *Chinese lexicography: A history from 1046 BC to AD 1911*. Oxford: Oxford University Press.

Yu, Bin 虞斌. 1974. "*Ping Xiandai hanyu cidian chongyinben*" 评《现代汉语词典》(重印本) [Comment on the reprint of the *Xiandai hanyu cidian*]. *Beijing daxue xuebao (zhexue shehui kexue ban)* 北京大学学报 (哲学社会科学版), 1974 (3): 91-94.

Zhang, Daosheng 张道生. 2016. "*Shixian cishu qiangguo zhanlüe mubiao de celüe tanxi*" 实现辞书强国战略目标的策略探析 [An analysis of the strategies to realize the strategic goal of a lexicographical power]. *Chuban faxing yanjiu* 出版发行研究, 2016 (6): 27-29.

Zhang, Zhiyi 张志毅, Zhang Qingyun 张庆云. 2015. *Lilun cidianxue* 理论词典学 [Theoretical lexicography]. Beijing: Shangwu yinshuguan.

Zhang, Zhiyi 张志毅. 2010. "*Cishu guancha: cishu qianguo yannjiu you duo yuan*" 书界观察: "辞书强国" 究竟有多远? [How far from the "lexicographical power"?]. *Renminwang* 人民网, October 12, 2010. Accessed January 10, 2020. http://culture.people.com.cn/GB/87423/12923586.html.

Zhao, Lianzhen 赵连振 and Geng Yundong 耿云冬. 2015. "*Liang ci cishu guihua yilai wo guo hanyu zhuanke cishu fazhan taishi shulüe*" 两次辞书规划以来我国汉语专科辞书发展态势述略 [Development trend of Chinese monolingual specialized lexicography since the publication of the two National Plans]. *Cishu yanjiu* 辞书研究, 2015 (2): 6-13.

Cited reference works

Chen, Zhli 陈至立 et al. 2020. *Cihai* 辞海 (*di qi ban* 第七版). Shanghai: Shanghai cishu chubanshe.

Cihai bianji weiyuan 辞海编辑委员 (eds). 1979. *Cihai* 辞海 (*san juan ben* 三卷本). Shanghai: Shanghai cishu chubanshe.

Hanyu da zidian weiyuanhui 汉语大字典委员会 (eds.). 1986-1989. *Hanyu da zidian* 汉语大字典 (Vols. 1-8). Wuhan: Hubei cishu chubanshe; Chengdu: Sichuan cishu chubanshe.

Li, Hangjian 李行建 et al. (eds.). 1989. *Xin ci xin yu cidian* 新词新语词典. Beijing: Yuwen chubanshe

Mei, Deming 梅德明 et al. (eds.) 2017. *Yuyanxue yu yingyong yuyanxue baike quanshu* 语言学与应用语言学百科全书. Beijing: Beijing daxue chubanshe.

Qian, Zhonglian 钱仲联 et al. (eds.). *Zhongguo wenxue da cidian* 中国文学大辞典. Shanghai: Shanghai cishu chubanshe.

Shang, Yongqing 尚永清 et al. (eds.). 1991. *Xin Han Ri cidian* 新汉日词典. Beijing: Shangwu yinshuguan; Tokyo: Riben xiaoxueguan.

Shanghai waiguoyu xueyuan上海外国语学院. 1977. *Han E cidian* 汉俄词典. Beijing: Shangwu yinsghuan.

Xin ying han cidian bianxiezu 新英汉词典编写组 (eds.). 1975. *Xin Ying han cidian* 新英汉词典. Shanghai: Shanghai renmin chubanshe.

Xinhua zidian 新华字典. 1979. Beijing: Shangwu yinshuguan.

Zhongguo shehui kexueyuan yuyan yanjiusuo cidian bianjishi 中国社会科学院语言研究所词典编辑 (eds.). 2016. *Xiandai hanyu cidian di qi ban* 现代汉语词典第7版. Beijing: Shangwu yinshuguan.

Chapter Two

Linguistic Features and Literary Value of *Almanca Tuhfe* / *Deutsches Geschenk* (1916)

Harald Bichlmeier
and Güler Doğan Averbek

Abstract

Versified dictionaries are bilingual/multilingual glossaries written in verse form to teach essential words in any foreign language. In the Islamic culture, versified dictionaries were produced to teach the Arabic language to the young generations of Muslim communities not native in Arabic. In the course of time, many bilingual/multilingual versified dictionaries were written in different languages throughout the Islamic world.

In this article, bilingual versified dictionaries having Turkish as one of their languages and a Western European language as the other one are compared with dictionaries written in the Balkan languages. The need for such dictionaries is also briefly discussed.

The main focus of our study is on the Turkish-German versified dictionary titled *Almanca Tuhfe / Deutsches Geschenk* [German Gift], published by Dr. Şerefeddin Pasha in 1916. This dictionary is the only dictionary in verse ever written that combines these two languages. We will show the characteristics and basic linguistic features of this unique dictionary as well as give an evaluation of its literary value as far as rhyme, meter and the basic attributes of versified dictionaries in Islamic culture are concerned. We describe the differences between this work and those in other languages of the non-Muslim communities of the time are described, and we also discuss the place of this dictionary among the bilingual/multilingual versified dictionaries having Turkish as one of their languages.

1. Versified glossaries

Versified glossaries are bilingual/multilingual glossaries written in verse form to teach essential words in any foreign language. In general, poetic books have been used for centuries in the tradition of Islamic education. Namely, education benefits from the use of well-rhymed and -metered texts that can easily be remembered. Such texts are aimed at students in order that they may easily acquire information, so that the information remains permanently in the students' memory. In Islamic culture, versified glossaries were produced to teach Arabic to young generations of Muslim communities not native in Arabic. Over time, many bilingual/multilingual versified dictionaries were written in different languages throughout the Islamic world.

The first known example of the bilingual versified glossary tradition is *Nisab al-Sibyan* (1221), which aimed at teaching Arabic words to persons speaking Persian (Fārsī) at that time. It was written by Farahi (d. 1242) in Afghanistan. In the Islamic tradition from the thirteenth until the twenty-first century, various versified glossaries have been written in many languages, namely in Arabic, Persian, Turkish, Hindi, Urdu, Kurdish, Sanskrit, Pashto, Tabari, English, German, French, Bulgarian, Greek, Armenian, Bosnian and Albanian, thus covering areas from Western Europe via the Balkans to the Indian subcontinent.[1]

---

[1] Cf. Doğan Averbek 2018a, 87. Some important versified dictionaries having Turkish as one of the languages are the following: **Turkish-Arabic**: Abdüllatîf b. Melek, *Lugat-i Ferişteoğlu* (for the text, see Muhtar 1993), Sünbülzâde Vehbî, *Nuhbe-i Vehbî* (for the text, see Yurtseven 2003), Mehmed b. Ahmed er-Rumî(?), *Sübha-i Sıbyân* (for the text, see Kılıç 2007, 29-71), Fedâyî Mehmed Dede, *Tuhfe-i Fedâyî* (for the text, see Fedayî 2019); **Turkish-Persian**: Hasan Rızâyî, *Kân-ı Me'ânî* (for the text, see Turan 2012, 2939-92), Hüsâm b. Hasan el-Konevî, *Tuhfe-i Hüsâmî ez Mültekât-ı Sâmî* (for the text, see Arslan 2016), Şâhidî İbrâhîm Dede, *Tuhfe-i Şâhidî* (for the text, see Verburg 1997, 5-87), Sünbülzâde Vehbî, *Tuhfe-i Vehbî* (for the text, see Vehbî 2012); **Turkish-Arabic-Persian:** Hâkî Mustafâ-yı Üsküdarî, *Menâzımü'l-Cevâhir* (for the text, see Arslan 2011), Osman Şâkir b. Mustafâ Bozokî (Şâkirî), *Müselles-nâme-i Şâkir* (for the text, see Kaya & Ayçiçeği 2019); **Turkish-Albanian**: Nazîm, *Der Beyân-ı Türkî ma'a Lisân-ı Arnabud* (for the text, see Rossi 1946, 219-46), Mahmûd, *Dürre-i Manzûme* (for a study, see Doğan Averbek 2018b, 223-42); **Turkish-Bosnian**: Muhammed Hevâ'î Üsküfî, *Makbûl-ı Ârif* (for the text, see Uskufi 2001); **Turkish-Armenian**: Ref'î Kâlâyî, *Manzûm Lugat-ı Ermeniyye* (for the text, see Dankoff, Kut & Weitenberg 1996); **Turkish-Greek**: Hanyalı Osmân Nûrî, *Lugat-ı Manzûme-i Nûriyye berây-ı Terceme-i Lisân-ı Rumiyye* (for the text, see Erik 1982); **Turkish-French**: Yûsuf Hâlis Efendi, *Miftâh-ı Lisân* (for the text, see Yûsuf Hâlis 2006).

1.1 Formal characteristics of versified glossaries

Versified glossaries were written in the *aruz* meter[2] according to tradition and are usually composed of three parts. *Muqaddima*, the first part, always starts with the *basmala* (which indicates the key formula of Islam, “In the name of God, the Clement, the Merciful”), then it is followed by the *hamdala* (the saying of the formula *al-ḥamdu li’llāh*, “Praise belongs to God”) and the *salwala* (a phrase which contains the salutation to the prophet of Islam, one of the most used versions is *Allāhumma ṣalli ʿalā Muḥammadin wa-ʿalā āli Muḥammadin* which means “O God, send blessings upon Muhammad and upon the Progeny of Muhammad”). In Islamic culture, starting from the ninth century, the *muqaddima* became an independent part and developed a stereotyped form.[3] Finally, there is the *sabab-i nazm* (i.e., the reason for writing the poem). In this part, the poet talks about many issues ranging from the sources explored to the problems in the previous versified glossaries while listing the reasons for writing the glossary. There he tells how the work is created, and he generally mentions the name of the glossary along with his name.[4] Generally, together with the reasons, here the poet also tries to explain his methodology and sometimes the numbers used in his work. Generally, they use the same numbers from one to nine to indicate the equivalents in the languages of the glossary.

The second part of the glossary consists of many staves. The number of the staves depends upon the poet. Depending on the style, *bahr*s (the meter groups of *aruz*), and the meters of the staves are independent of each other. Resembling Farahi’s style as the pioneer of the glossary form in Persian, the meter of each stave is repeated in a line, and sometimes the remaining part of the couplet indicates the *bahr* and technical characteristics of the meter. Usually, one, or sometimes two couplets which come immediately before or after the meter couplet generally include some advice. This couplet contains statements as that of Farahi’s bards, and in glossaries they reflect an account containing a piece of advice, generally a

---

[2] This metrical system had been developed by the Arabs since pre-Islamic times and was systematized by el-Khalil b. Ahmad. It is widely used throughout the Islamic world up to the 20th century. For further information, cf. Çetin 1991, 424-37.
[3] For detailed information, cf. Freimark 1960-2002: 495-96; Durmuş 2006, 115-17.
[4] Cf. Doğan Averbek 2019, 69-70. As one of the remarkable names in the *tuhfe* tradition with his Turkish-Arabic, Turkish-Persian and Turkish-Arabic-Persian *glossaries*, Osman Şakir provides valuable information in his *muqaddima*s about the methods he applied in his works while criticizing the others. For details, see Özkan 2013, 428-62.

religious one. The educational mission of the glossary genre is strengthened in this way.[5] The main objective of this practice is to overcome monotony.[6]

The last part, which is called *khatima* ('epilogue'), informs the reader about the end of the dictionary.[7] In this section, the name of the dictionary is mentioned again, and the book ends with prayers and dates.[8]

In versified glossaries, while the first and third parts are written in the *maṯnawi* form (a poem written in rhyming couplets), the second part is generally written in *qasida* form.[9]

1.2 On the use and aims of versified glossaries in Ottoman times

An important issue with these glossaries is their use as a tool in education. These books were used as course books, especially in elementary schools. Although the main objective of such glossaries is the education of small children, the poets also wished that their glossaries be used by adults as well. There are examples of other versified glossaries that were written with different intentions. Primarily these are texts which do not include Arabic or Persian. In the last chapter of *Tuhfetü'l-Uşşâk* (1815),[10] a Turkish-Greek glossary in verse, Ahmed Fevzî (nineteenth century) wrote that his work is for fame and the fun of his friends:

---

[5] Cf. Doğan Averbek 2019, 71.
[6] Cf. Aksoy 1959, 218.
[7] These books were written as coursebooks, e.g., for children or for people who were willing to learn foreign languages. Although there is no such distinction like dictionary or glossary in the Ottoman tradition, according to their functions, these are glossaries. In the Ottoman tradition, these kinds of texts were called "tuhfe"s ("tuhfe" means "gift"). In its terminology, "tuhfe" refers to "versified glossary." For detailed information, see Doğan Averbek 2019.
[8] Cf. Doğan Averbek 2019, 71-2.
[9] This issue has not been clarified yet. Some scholars consider the number of couplets and conclude that the staves are written in *ghazal* or *qasida*, while other researchers concentrate on the rhymes only. However, looking at the history of the *qasida* form in Arabic poetry, it seems that the number of couplets is not a determining issue. Consequently, it is a point to qualify the poems with the aa, xa, xa rhymes which are frequently applied in such staves as the *qasida* form (cf. Doğan Averbek 2019, 69). For further information on qasida, cf. Elmalı 2001, 562-64.
[10] For a list of Turkish-Greek versified dictionaries, see Doğan Averbek 2018a, 110-11. The text of this one has not yet been studied entirely, but there are three different studies on it by same scholar: cf. Ölker 2004; Ölker 2005; Ölker 2009, 856-72.

Alup hâme ele Fevzî didüm târîhini böyle
Murâdum bu bula şöhret ola yârâna eglence[11]
["O Fevzî, I took the pen in my hand and wrote its date; my desire is that this book shall catch on and become an amusement for friends!"]

In *Makbûl-i 'Ârif* (1631-1632),[12] which is known to be the only Turkish-Bosnian versified glossary, Muhammed Hevâ'î (1601-1651) talks about his desire to write a booklet which has never been thought of.[13] In *Dürre-i Manzûme* (1835),[14] a Turkish-Albanian versified glossary, the poet Mahmud (according to the dictionary, probably an Albanian soldier of the Ottoman army, nineteenth century) states in the part where he explains the reasons for writing the work that Ottoman soldiers serving in Shkodra/Skutari did not speak Albanian, whereas the local folk did not speak Turkish.

Undoubtedly, during the Ottoman era, the language selection for versified glossaries is directly related to cultural interaction with its influences and current trends. For this reason, changes in the parties of cultural communication—for instance, as influenced by contact between countries and cultures as they change over time—can be traced in such glossaries. According to Yûsuf Hâlis Efendi (d. 1882), the poet of the Turkish-French glossary in verse, French is also *en vogue* (see Fig. 2-1 below). It became an obligation to study foreign sciences, and as the language of education was French, learning this language became required.[15]

---

[11] Cf. Ölker 2009, 861.
[12] Because it is considered as the first dictionary which uses the term "Bosnian," this versified dictionary is the most studied one that has Turkish as one of the languages. It is also worth mentioning that it is the first studied text among these kind of texts (cf. Blau 1868). Up to now, there are more than 25 works on it (for a bibliographic list, see Doğan Averbek 2018a, 99-101.
[13] Cf. Uskufi 2001, [np].
[14] For a detailed list of Turkish-Albanian versified dictionaries, see Doğan Averbek 2018b, 226-27.
[15] Güleç & Doğan Averbek 2019, 68.

Fig. 2-1: Turkish-French versified dictionary (*Miftâh-ı Lisân*, 1266/1849-1850).

2. Why did the Ottomans need a Turkish-German versified glossary?

Turkish-German relations were officially established more than 250 years ago.[16] The links between the German socio-political culture and the Ottoman Empire became official in 1878 and they developed in many fields

[16] Cf. Alkan 2015, 35.

of politics, military aspects, and the spheres of economics and culture continuing uninterruptedly until the end of World War I.[17]

It has been said that cultural opportunities are used to create a lasting effect on the Ottoman Empire.[18] "Germany applied cultural influences on the Ottoman country by sending numerous missionaries under the professions like physicians, teachers, clerics, academicians," and also tried to reinforce this influence by opening schools in different parts of the Ottoman territory.[19]

3. On *aljamiado* literature

Actually, we see here—in the German words of the lexicon—a specimen of the so-called *aljamiado* literature, i.e., literature (in the broader sense of the word) written in Arabic script in a language for which the Arabic script is not typically used. This generally accepted and widely known *terminus technicus* is based on the Spanish word *aljamiado*, which itself is derived from Spanish *aljamía*, a loanword from Arabic *al-ʼaʽǧamiya* 'the non-Arabic, the foreign one.'[20]

The use of Arabic script for languages regularly using Latin or Cyrillic script was quite wide-spread during a certain period. It was mostly employed by the Islamicized populations that had come under Arab or later Ottoman rule: e.g., speakers of Spanish in the Middle Ages, speakers of the Bosnian variant within the Serbo-Croatian dialectal continuum from the sixteenth century onwards up to the twentieth century and speakers of Belarusian from the seventeenth to eighteenth centuries onwards until the middle of the twentieth century. The case of Belarusian slightly differs from the other two, as the Islamicized Turkish-speaking Tatars had switched to Belarusian after they had been transferred from the Crimea region to the northern parts of the Lithuanian-Polish Commonwealth. Bosnian texts were even printed in Arabic script since the end of the nineteenth century in Bosnia and Hercegovina, occupied by Austria-Hungary since 1878 (Austria-Hungary thus supported the building of a Bosnian-Muslim identity to form a counterweight against Croatian and Serbian nationalism),[21] and after 1918 in Yugoslavia. This printing tradition ended after the German

[17] Cf. Reyhan 2005, 7, 11.
[18] Cf. Reyhan 2005, 32.
[19] Cf. Reyhan 2005, 38: "Almanya, Osmanlı ülkesi üzerinde kültürel nüfuzunu, tabib, öğretmen, din adamı, akademisyen vb. görünümünde sayısız misyoner göndererek kuruyordu."
[20] Cf. Kontzi 1980; Hadžijahić 1941.
[21] Cf. Neweklovsky 1996, 63-67.

invasion of Yugoslavia in April 1941. Attempts at a revival of writing in the so-called *Arebica*[22] after the breakup of Yugoslavia (1991-1992) and the civil war (1992-1995) have not led to anything worth mentioning so far. Even in Bosnia, the concept of writing Bosnian with Arabic script is mostly regarded as a somewhat weird idea even by the larger parts of the Muslim community. Belarusian texts in Arabic were printed in Poland before World War II, but with the war, this tradition ended, just like in Bosnia. The Warsaw city council had even agreed upon building a mosque for the Tatars—but this decision was not made before the spring of 1939, and thus the mosque was never built.

4. Almanca Tuhfe

4.1. General data

The glossary *Almanca Tuhfe/Deutsches Geschenk* published in Istanbul in 1916 is of particular importance because it also shows with respect to the humanities that the Ottoman-German relations were at their peak. Writing this versified glossary seemed vital as it points to the need to know German, as stated in the prose introduction. Indeed, Şerefeddin Pasha, the author, explains that because Ottoman-German relations are growing, the two nations should know each other's languages.

We do not have much information about Şerefeddin Pasha. He was a military physician in the Ottoman Army and served as the Director of Van Sanitary before 1916. He also served at the military hospital in Zeytinburnu in 1908.[23] According to a newspaper record he died in 1926.[24]

The booklet has 52 pages (see Fig. 2-2 below). Although the author says in the introduction that he has used more than 2000 words, the work contains German equivalents of 1162 Turkish words and phrases. The author may have meant the total number of words both in German and Turkish.

[22] Cf. Hadžijahić 1955.
[23] See Yıldırım 2006, 347. There is also an archival document which proves his employment at this hospital (State Archives of the Turkish Republic, 253-257, Y.PRK.ASK).
[24] Cf. *Haber* 1926, 4.


آلمانجه تحفه

محرری

دوقتور شرف الدین پاشا

Deutsches Geschenk

Verfasser

Dr Scherefeddin Pascha

هر حق محفوظدر : تقلید ایدنلر نظاماً مسؤلدر

CONSTANTINOPEL
Buchdruckerei
von
Osman Bäi
Tchemberli-Tasch
1916

استانبول
مطبعهٔ عثمانیه
چنبرلی‌طاش
١٣٣٢

صاحبی طرفندن مختوم اولمیان نسخه‌لر ساخته‌در

Fig 2-2: Title page of *Almanca tuhfe.*

The author stated that he learned German in private lessons and by personal effort because he anticipated this need:[25]

> Devlet-i ᶜAliyye-i Os̲māniyye ile onun muḥibb-i ḫālīsi ve müttefiḳ-i ṣādıḳı olan Almanya devlet-i feḫīmesi beyninde her türlü münāsebet-i düveliye günbegün artmaḳda bulunduġundan ᶜumūmen ᶜOs̲mānlı ve Cermen aḳvām-ı necībesi yekdīgerinin lisānlarını bilmege mecbūriyet ḥāṣıl olmaḳdadır. Binā'enᶜaleyh ᶜOs̲mānlılar Almancayı, Almanlar daḫi

[25] Şerefeddin Pasha 1916, 3-4.

> müteḳābilen Türkçeyi ögrenmeleri lüzūmına mebnī Berlin mekteblerinde Türkçe ve İstanbul mekteblerinde Almanca lisān dersleri ᶜilāve edilmekde bulunduġu maᶜa'l-mesār görülmekde ve işitilmekdedir.
> İşte bu lüzūmu çākerleri daha evvelce ḥissederek Van vilāyeti ṣıḥḥiye müdīriyetinde bulunduġum zamān esātiẕe-i kirāmımdan fisyolojist merḥūm Şākir Paşa'nıñ maḫdūmu doḳtor operatör ᶜĀkif Bey'den az bir müddet Almanca dersi alaraḳ baᶜdemā kendi kendime taḥṣīle devām etmekdeyim.[26]

4.2. Structure and contents of **Almanca tuhfe**

Beginning with a prose preamble, *Almanca Tuhfe/Deutsches Geschenk* is composed of 23 sections, consisting of a *sabab-i nazm* part, 21 sections forming the glossary part and an epilogue of two stanzas. The work consists of three stanzas and 295 couplets. This versified dictionary has been published in the meantime. The publication (Doğan Averbek and Bichlmeier 2020) includes the facsimile and the transliteration of the whole text with a commentary as well as an introduction on versified dictionaries as coursebooks.

The glossary part begins with the words *Allah* and *prophet* following the practice seen in versified glossaries. In addition, the German form of the *basmala* is given in the work, which is not common in such glossaries (see Fig. 2-3 below).[27]

p. 5, § 3 im namen ġotes (aim namn ġutas) des albarmherziġin des erbarmenden (ds ālbarmhrziġin[28] ds ᵓarbarmndn)
Im Namen gottes des allbarmherzigen des Erbarmenden

---

[26] Translation: "The relations between the Ottoman State and Germany, which is a loyal ally to the Ottoman State, are increasing day by day. Therefore, the Ottoman and German nations need to know each other's languages. For this reason, depending on the needs of the Ottomans and the Germans, Turkish language courses in schools located in Berlin and German language courses in schools located in Istanbul have started. As I felt this necessity before, I took private German lessons from Operator Doctor Akif Bey, the son of my esteemed teacher, physiologist Şakir Pasha, while I was Director of Van Sanitary and then continued to learn German on my own."

[27] This study transcribes the German words written in Arabic letters (giving the probably intended Ottoman interpretation) with the transliterations given in brackets as well as Turkish text.

[28] <ālbarmhrziġin> is wrongly spelled for *<ālbarmhrčiġin>; probably the German spelling <z> [ʦ] was confused with the spelling of Arabic <z> [z].

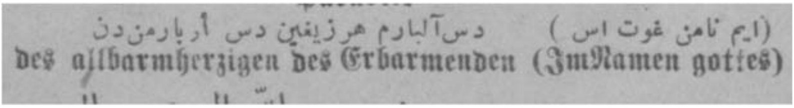

دس آلبارم هرزیغین دس أربارمن دن (ایم نامن غوت اس)
des allbarmherzigen des Erbarmenden (ImNamen gottes)

Fig. 2-3: *Basmala*, *Almanca tuhfe*, p. 5.

As a preference not found in other versified glossaries, sections that make up the glossary chapter are in this case classified according to their subjects. The section titles are arranged accordingly. The author notes in the introduction to the glossary that this classification allows him to gather related words together:

> p. 5 Tuḥfe-i Vehbī'ye beñzer Alman luġāti yazdımmütecānis
> olanları hep bir araya ṭopladım
> [I have written a German dictionary which resembles 'Tuhfe-i Vehbî' and I gathered similar words together.]

The titles of the sections that make up the glossary part and their respective number of couplets are as follows (Table 2-1):

	Turkish Title	**German Title**	**Meaning**	**Number of Couplets**
1.	*Kâ'inât*	*Die Wesen*	The beings	11
2.	*E'âzım*	*Die Grossen*	The 'big ones'	7
3.	*Ebeveyn hısım ve akrabâ*	*Die Eltern*	The parents	7
4.	*Sayılar*	*Die Zahlen*	The numbers	17
5.	*Hafta Eyyâmı*	*Die Wochen Tage*	The days of the week	3
6.	*Senenin on iki ayları*	*Die Monate Des Jahres*	The months of the year	4
7.	*Sıfât-ı muhtelife*	*Das Bei Wort*	The adjective	17
8.	*Me'kûlât ve meşrûbât*	*Die Speise Und Getränke*	Food and drinks	15
9.	*Yemişler ve sebzeler*	*Die Früchte Und Gemüse*	Fruits and vegetables	13

	Turkish Title	German Title	Meaning	Number of Couplets
10.	*Hayvânât – kuşlar – balıklar*	*Die Thiere Und Vögel Und Fische*	Animals and birds and fish	15
11.	*A'zâ-yı beden*	*der Menschliche Körper*	The human body	13
12.	*Melbûsât*	*Die Kleidungen*	Clothing	10
13.	*Mesken – ev eşyâsı*	*Wohnung-Möbel*	Flat and furniture	18
14.	*Almanca fi'ller*	*Das Zeit Wort*	The verb	51
15.	*Mekteb*	*Die Schule*	The school	15
16.	*Lugât-i muhtelife*	-		18
17.	*Harf-i cerler*	*Das Vorwort*	The preposition	12
18.	*Zamîrler*	*Die Pronomen*	The pronoun	15
19.	*Meşhûr şehrler*	*Die Städte*	The cities	6
20.	*Mükâleme*	*Die Spräche*	The languages	24
21.	*Sâ'atlere dâ'ir*	*Von der Uhr*	About the clock	4

Table 2-1: Structure of *Almanca tuhfe*

The explanations are in Turkish. Şerefeddin Pasha took into consideration the needs to know each other's languages which arose from the development of the relations between the Ottoman State and Germany. He writes that he tried to write a work similar to the famous Turkish-Persian versified glossary titled *Tuhfe-i Vehbî*:

p. 5 Tuḥfe-i Vehbī'ye beñzer Alman luġāti yazdım.
[I have written a German dictionary which resembles 'Tuhfe-i Vehbî.']

The author's purpose for his work, namely, learning the vocabulary of the language through memorization, is suggested by the following verse:

p. 17, § 79 bu luġatleri ḥıfẓ etmek hem elzemdir hem elbetdir
[It is necessary to memorize these words.]

Şerefeddin Pasha summarizes the reasons for writing his work in his prose introduction titled “Başlangıç,” from which already some critical points were mentioned above:

> Devlet-i ᶜAliyye-i O̱smāniyye ile onun muḥibb-i ḫālīsi ve müttefiḳ-i ṣādıḳı olan Almanya devlet-i feḫīmesi beyninde her türlü münāsebet-i düveliye günbegün artmaḳda bulunduġundan ᶜumūmen ᶜO̱smānlı ve Cermen aḳvām-ı necībesi yekdīgerinin lisānlarını bilmege mecbūriyet ḥāṣıl olmaḳdadır. Binā’enᶜaleyh ᶜO̱smānlılar Almancayı, Almanlar daḫi müteḳābilen Türkçeyi ögrenmeleri lüzūmuna mebnī Berlin mekteblerinde Türkçe ve İstanbul mekteblerinde Almanca lisān dersleri ᶜilāve edilmekde bulunduğu maᶜa’l-mesār görülmekde ve işitilmekdedir.
>
> İşte bu lüzūmu çākerleri daha evvelce ḥissederek Van vilāyeti ṣıḥḥiye müdīriyetinde bulunduġum zamān esātiẕe-i kirāmımdan fisyolojist merḥūm Şākir Paşa’nıñ maḫdūmu doḳtor operatör ᶜĀkif Bey’den az bir müddet Almanca dersi alaraḳ baᶜdemā kendi kendime taḥṣīle devām etmekdeyim.
>
> Almancayı mükemmelen ögrenmeye benim gibi heveskār olan iḫvān-ı kirāma ve henüz taḥṣīle başlamış bulunan evlād-ı vaṭana ḳolayca luġat belletmek için Fārisīce *Tuḥfe-i Vehbī*’ye benzeyen *Almanca Tuḥfe* nāmıyla ᶜācizāne ve nāçīzāne bir luġatçe yazdım. Derūnuna iki bini mütecāviz en lüzūmlu luġāti derc ve cemᶜ ederek ḳāri’īn-i kirām ḥażerātına hediyyeten taḳdīm eyledim.
>
> Faḳaṭ bundan maḳṣadım ḥāşā Sünbülzāde merḥūmuñ tuḥfesine manẓūm naẓīre yapacaḳ ḳadar liyāḳatim var demek degil belki bugün bildigimiz ve söyledigimiz Fārisī luġatlerini vaḳtiyle hepimize ögreten ve yüz küsūr sene evvel te’līf edilmiş olan ġāyet nāfiᶜ böyle bir kitābıñ uṣūl-i taḥrīrini taḳdīren taḳlīd ederek hem mü’ellifiniñ rūḥunu şād etmek ve hem de ṭarz-ı mezkūr üzere daha nāfiᶜ ve mükemmel diger es̱erler meydāna getirilmesiyçin erbāb-ı fen ve iḳtidārı ve lisān-āşināyān-ı zamānı min-ġayri-ḥad teşvīḳ eylemekdir.
>
> Bālāda ᶜarż ve beyān edildigi vechile Alman lisānının tekmīl ġavāmıżına henüz vuḳūf-ı tām peydā edemedigimden nāşī tuḥfemdeki luġāt-i münteḫaba Ticāret ve Zirāᶜat Neẓāreti Orman Müşāvir-i ᶜĀlīsi Fayt Bey’in tercümān-ı ẕī-iḳtidārı ve Orman Mektebi Almanca Muᶜallimi Mübāhāt Bey’in naẓar-ı tedḳīḳlerine ᶜarż edildikde luṭfen taṣḥīḥ ve taḳdīr buyurulmuş olması üzerine es̱er-i ᶜācizinin her türlü nevāḳıṣıyla berāber enẓār-ı ᶜāmmeye važᶜına cesāret geldi. Artıḳ ḳāri’īn-i kirām ḥażerātı ṭarafından görülecek her nevᶜ sehv ve ḫaṭāları̃ñ daḫi ᶜafv

> buyurulacağını ve bu bābdaki ḥüsn-i niyetime bağışlanacağını ḳaviyyen ümīd eylerim. Ve mina'llāhi't-tevfīḳ. [29]

As it is understood, the author thought that the versified glossary genre, which had already made significant contributions to foreign language teaching, would also be useful in learning German.

One of the striking points here is the desire of the author to shed light on a new path leading to the composition of such works. In the prose introduction, he argues that although there are many errors in the work, he has paved the way for composing the Turkish-German versified dictionary. In one couplet, the author explicitly states the necessity of knowing Turkish and German:

p. 41, § 231 türkiş Türkçe ᶜOs̱mānlıca doyç Almanca dimek işbu
lisānları bundan ṣoñra herkes bilmek gerek
['türkiş' means Ottoman Turkish and 'doyç' means German, from now on everyone should know these languages.]

4.3 Metrics

The *aruz* meter, which was invented by Arabs and systematized by al-Khalil, was adopted to Turkish poetry starting from the eleventh century.

[29] Translation: "The relations between the Ottoman State and Germany, which is a loyal ally to the Ottoman State, are increasing day by day. Therefore, the Ottoman and German nations need to know each other's languages. For this reason, depending on the needs of the Ottomans and the Germans, Turkish language courses in schools located in Berlin and German language courses in schools located in Istanbul have started. As I felt this necessity before, I took private German lessons from Operator Doctor Akif Bey, the son of my esteemed teacher, physiologist Şakir Pasha, while I was Director of Van Sanitary and then continued to learn German on my own. I wrote a glossary called *Almanca Tuhfe* (German Gift) similar to Persian *Tuhfe-i Vehbî*, to make it easier to memorize words for those who want to learn German like me and for the students. In this glossary, I used more than 2000 words. However, my intention is not to make a versified parallel of Vehbî's work. My purpose is to encourage the arts and language masters to carry on the spirit of the author as well as to write more useful and excellent works in this way by imitating such a useful work written more than one hundred years ago that taught us all Persian words we know. Before publishing the book, I have shown it to the Counselor of the Ministry of Commerce and Forestry, Mr. Veit, and the German Lessons Teacher of the School of Forestry, Mr. Mübahat, to correct and control the words I used in the book because I am not yet fully qualified. Now I publish this book and ask the reader to forgive my mistakes."

This meter is based on the long and short vowels of the syllables in every line in any poem.[30] The syllabic meter is the national one that is used in Turkish folklore poetry during the Islamic era. According to this meter, every line has the same number of syllables, from 7 to 15. It is called *hece* ['syllabic'] or *parmak hesabı* ['finger counting'] in Turkish.[31] It is possible to apply both metric systems to the following couplet of Yunus Emre (fourteenth century):

> Aşkın aldı benden beni bana seni gerek seni
> Ben yanaram düni güni bana seni gerek seni
> [Your love has driven me away; I need only you,
> I am suffering every moment; I need only you!]

While this couplet was written with 8+8 = 16 syllables according to the syllabic meter, it is also in compliance with "mustafʿilun mustafʿilun mustafʿilun mustafʿilun" [i.e. dum-dum-da-dum dum-dum-da-dum dum-dum-da-dum dum-dum-da-dum] meter of the Rajaz *bahr* of *aruz*.

On the other hand, *Almanca tuhfe* is a remarkable example of poetry. Although it was written within the versified glossary tradition, some problems related to the meter used in the work must be mentioned. Although the author says that he used the meter "faʿilatun faʿilatun" (i.e. dum-da-dum-dum dum-da-dum-dum) of *aruz*, it is impossible to apply the given meter to the work. However, it is not reasonable to imagine that an educated man of that time did not know the *aruz* meter. An elementary school education of that time would have been certainly able to teach him this. Besides, in his work, the author confesses that he learned and memorized the work titled *Tuhfe-i Vehbî,* which is also useful for learning *aruz*. So, it cannot straightforwardly be understood why the poet said that he used this meter. If one evaluates the work from the metrical point of view, it is possible to conclude that it was written in the syllabic meter instead of *aruz*. However, we cannot say that the poet has always remained faithful to this meter either. Nevertheless, if we accept that the syllabic meter was used, this glossary would be, as far as we know, the first one written in the syllabic meter among the versified glossaries. Concerning metrics, further studies will be necessary, however.

[30] For detailed information, cf. Gibb 1900-09, 1, 104-10.
[31] For detailed information, cf. Gibb 1900-09, 1, 104-10.

4.4. On the lexicon

In some lines, there are—following the versified glossary tradition—didactic comments instead of vocabulary. The author did not give German equivalents for these comments:

p. 6, § 8 cān-ı ᶜazīzi besleyüp fikri güzel parlatmalı
[One must feed the supreme soul and brighten the idea properly.]
p. 6, § 9 kā'ināta çeşm-i ᶜibretle bir kerecik baḳmalı
[One must draw a lesson from the cosmos.]

Sometimes, the author also gives extra information about the word he used or some quotation from other literature. The following couplet exemplifies this situation:

p. 22, § 107 pfav ṭāvus ḳuşınıñ da tügleri pek reng-āmīzdir
ayaḳları pek çirkīndir anuñ-çün feryād-engīzdir
[The colors of the peacock / 'Pfau' are gorgeous,
but its feet are hideous. That is why it screams.]

In the glossary, while the equivalents of the German words were given, Arabic and Persian words, which are widely used in Turkish, were also given. Sometimes we can observe that the same word is given additionally in Arabic or Persian to its Turkish equivalent. For example, the following couplet includes the Arabic equivalent of the Turkish word *baş* and the Persian equivalent of the Turkish word *göz*:

p. 22, § 110 şedel ḳafa ṭası dimekdir ḳopf re'sdir hem başdır
das avge göz hem çeşmdir avġın bravue ḳaşdır
['Schädel' means skull, 'Kopf' means head,
'das Auge' means eye and 'çeşm' in Persian, 'Augenbraue' means eyebrow.]

And in the following line, the Greek word *prostela* (*apron* in English) was used with the Turkish word *önlük*:

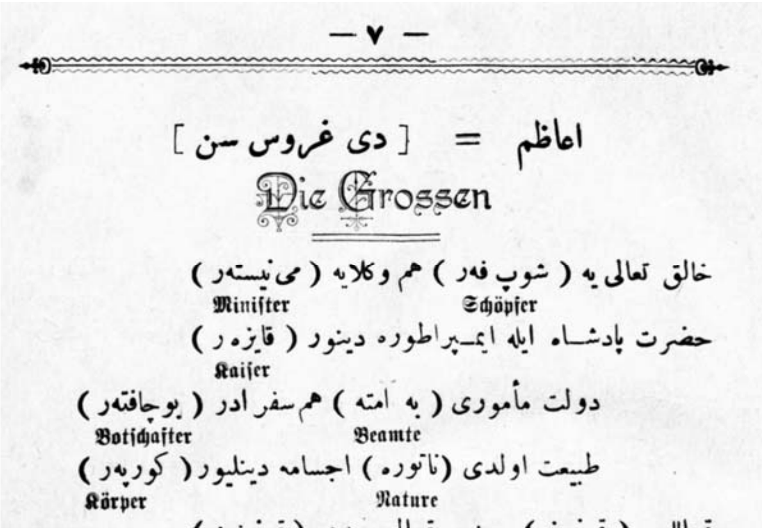
p. 25, § 130 öñlük prostela şürçe naᶜleyn daḫi holç pantoffel
[Apron and 'prostela' mean 'Schürze', clog means 'Holzpantoffel.']

The sentence patterns given under the title of "Mükâleme," which means "Dialogue", are also worth mentioning. This section, which contains about

60 sentences and their equivalents, should be carefully considered in terms of versified glossaries. We have not seen another example like this collection of sentences and phrases in any other versified glossary we consulted.

5. Linguistic features of the German words

In the glossary, the German words are given as integrated forms in the Ottoman text; the original German form is given below in Fig. 2-4 in Gothic letters. Cf. the beginning of p. 7:


— ٧ —

اعاظم = [دی غروس سن]

Die Grossen

خالق تعالی یه (شوپ فەر) هم وکلایه (می نیستەر)

Minister Schöpfer

حضرت پادشاه ایله ایمپراطوره دینور (قایزه ر)

Kaiser

دولت مأموری (به امته) هم سفرادر (بوچافتەر)

Botschafter Beamte

طبیعت اولدی (ناتوره) اجسامه دینلیور (کورپەر)

Körper Nature

Fig. 2-4: *Almanca tuhfe*, p. 7.

The glossary seems to have been produced rather quickly: there are dozens of typos in the German words; initial capital letters at the beginning of names and nouns are not used consistently; on the other hand, sometimes also other kinds of words (e.g. adjectives) are written with initial capitals; and finally, compounds are sometimes written as separate words. See, for instance, *Kammer präsident* vs. *Feld-marschal* vs. *Hauptmann* (p. 7; see Fig. 2-5 below):

p. 7, § 17 *ḳammerprezident* (qāmer prezidnt)[32] *Meclis-i Mebᶜūs̱ān re'īsidür*
Kammer präsident
ordu müşīri feldmarşal (feldmaršal) *paşalar hem general* (general) *dir*
Feld-marschal[33] General

p. 7, § 18 *mīr-alaya uberst* (auberst) *dinür mülāzıma liyutinan* (liutynan)[34]
Oberst Lieutenant[35]

biñbaşıya mayör (mayur) *dinür hem yüzbaşıya havuptman* (hauuptman)
Major Hauptmann

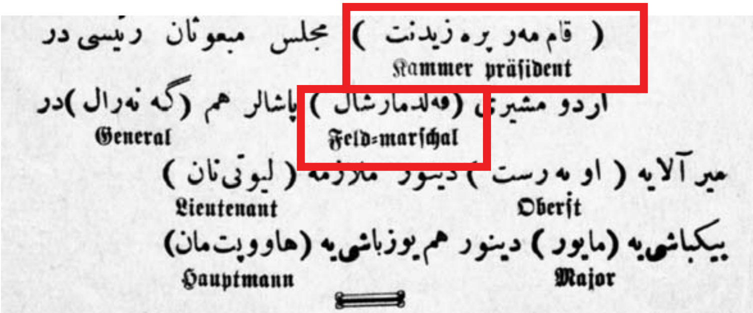

(قام‌مه‌ر پره‌زیدنت) مجلس مبعوثان رئیسی در
Kammer präsident
اردو مشیری (فه‌لدمارشال) پاشالر هم (گه‌نه‌رال)در
Feld-marschal General
میر آلایه (او به‌رست) دینور ملازمه (لیوتی‌نان)
Oberst Lieutenant
بیکباشی‌یه (مایور) دینور هم یوزباشی‌یه (هاوویت‌مان)
Major Hauptmann

Fig. 2-5: *Almanca tuhfe*, p. 7 continued.

As mentioned above, compounds are usually also written in two parts. But this might have also been done in the readers'—i.e. in the learners'—interest to make it clear of what parts a particular compound is made. But nothing about wanting to do so is mentioned by Şerefeddin Pasha in the introduction. So, for the moment, this is mere guesswork.

Thus, the really interesting point about this glossary is that German words and sentences are rendered in Arabic script by a writer who was trained in the Ottoman orthography of his times. But Şerefeddin Pasha, the author of the 1916 glossary *Almanca tuhfe*, used a way of writing the German words which lagged clearly behind the possibilities of the Arabic

[32] The transliteration renders the geminate of the Standard German spelling.
[33] Modern spelling: *Feldmarschall*.
[34] The transliteration reflects the older/traditional pronunciation borrowed from French: [ljøt(ə)nã]; modern German pronunciation: [ˈlɔɪtʰnantʰ].
[35] Modern spelling: *Leutnant*.

script (which, e.g., has diacritics on the one hand and signs for long vowels on the other hand, which might enable a writer to constantly and consequently differentiate long and short vowels) and even the standards used by the Ottomans for writing Turkish at the time.

Surprisingly, Şerefeddin Pasha does not employ another orthographical convention of Ottoman spelling either. The Arabic language has plain consonants (*t*, *k*, *s*, *d*) and the so-called emphatic consonants (*ṭ*, *q*, *ṣ*, *ḍ*, pronounced with the tongue farther back in the mouth than in the case of the plain consonants), which are represented in the Arabic alphabet by different signs. In the Turkish phoneme system, this differentiation does not exist, so these consonant signs are free to be used for something else, such as marking certain vowel differences. The signs for plain consonants are written when the following vowel is a front vowel and the signs for emphatic consonants when the following vowel is a back vowel.

Not making use of this possibility of marking differences in vowel length etc. seems the more surprising, as the language form Şerefeddin Pasha gives in Arabic script is the standard German variety of the time, for which at that time already existed dictionaries giving the appropriate pronunciation of German words (cf. e.g., Siebs 1905, 1915). So, there would have been the chance to develop a proper and straightforward system of spelling, taking into account all phonologically relevant features. Of course, we do not know whether Şerefeddin Pasha knew of such books or had access to them.

The only task of Şerefeddin Pasha's would have been to make up a list of correspondences between the approximately 40 phonemes of Standard German and their Arabic equivalents. However, there is not even the slightest attempt at making a distinction between short and long vowels; they are written identically. Also, the quality of the vowels is not observed consistently—as can be seen in the case of the variety of phonemes represented by the letter *wāw* (cf. Table 2-2). This somewhat deficient way of writing sometimes even leads to such homographs as <ġut> *gut* /gūt/ [guːt^{h}] 'good' vs. <ġut> *Gott* /got/ [gɔt^{h}] 'God.'

و	u	/u/ (in inlaut) <mutter> *Mutter* 'mother', <hund> *Hund* 'dog' /ū/ (in inlaut) <hun> *Huhn* 'hen' /o/ (in inlaut) <tuḫtr> *Tochter* 'daughter', <ārtišuqke> *Artischocke* 'artichoque'

		/ō/ (in inlaut) <zun> *Sohn* ‘son’, <huze> *Hose* ‘trousers’, <huḫ> *hoch* ‘high’
		/ü/ (in inlaut) <muqke> *Mücke* ‘mosquito’
		/ü/ (in anlaut) <ubersessin> *übersetzen* ‘translate’
		/ǖ/ (in inlaut) <zus> *süß* ,sweat‘, <tur> *Tür* ‘door’
		/ö/ (in inlaut) <šupfer> *Schöpfer* ‘creator’
		/ȫ/ (in inlaut) <fugel> *Vögel* ‘bird’, <šun> *schön* ‘beautiful’
	w	/v/ (in anlaut) <wasser> *Wasser* ‘water’

Table 2-2: Sounds rendered by the letter *waw.*

To a certain degree, this inconsistency is due to the deficiencies of the Arabic alphabet as far as vowels are concerned. Classical Arabic had a straightforward triangular vowel system of *a*, *i*, *u* and *ā*, *ī*, *ū*. While the pre-classical Arabic alphabet had only signs for the long vowels, with the letters *ālef* for *ā*, *yā* for *ī* and *y* and *wāw* for *ū* and *w*, the classical Arabic alphabet developed three diacritic signs for marking the short vowels and another one for marking reduplicated consonants. In the Persian variant, a special sign for /e/ arose as well, originally a letter for /h/. This letter is used in both ways in the 1916 dictionary (cf. Table 2-3):

ه	h	/h/ <huze> *Hose* ‘trousers’, <huḫ> *hoch* ‘high’
	e	/e/ [ɛ] (in inlaut) <šperber> *Sperber* [ˈʃp^{h}ɛrbɐ], [ˈʃp^{h}ɛrbɛʁ] ‘sparrow hawk’, <fet> *fett* ‘fat’
		/e/ [ə] (in unaccented syllables, esp. in auslaut) <qawen> *kauen* ‘chew’, <tawen> *tauen* ‘thaw’
		/ē/ [eː] (in inlaut) <qamel> *Kameel* (Modern Germ. *Kamel*) [k^{h}aˈmeːl] ‘camel’, <leber> *Leber* [ˈleːbɐ] ‘liver’, <kele> *Kehle* [ˈk^{h}eːlə] ‘throat’
		/ē/ [eː] (in auslaut) <fe> *Fee* [feː] ‘fairy’
		/ǟ/ [ɛː] <šedel> *Schädel* [ˈʃɛːdəl] ‘scull’

Table 2-3: The values of <h> in the 1916 dictionary.

Consonants have a somewhat better fate, as the consonant system of German is rather simple, and the Arabic alphabet in its Persian and Ottoman variants had a big enough inventory of letters available for representing

German consonantal phonemes. Still, even in this case the spelling is neither completely phonetic nor completely phonologic. For example, the initial antevocalic German <s-> /z-/ is mostly rendered as <z-> in the Arabic script in the dictionary, but there are four loanwords (from Romance languages) which are written with <s->: <salāt> *Salat* ‘salad’, <sardine> *Sardine* ‘sardine’, <sufa> *Sofa* ‘sofa’, <suldat> *Soldat* ‘soldier.’ There is no reason for doing so; we might surmise that in these four words the German spelling was imitated, while in all the others the German pronunciation was rendered correctly.

On the other hand, we can observe a tendency to reflect the spelling of the German words represented in the Arabic transcription. For instance, word-internal written geminates appear as geminates in the Arabic transcription as well (NB: there are no phonological or phonetic geminates in German nowadays). Meanwhile, word-final written geminates are represented by a single grapheme—in accordance with phonology (and phonetics).

Compare the following examples:

> <au**ff**yčir> *Offizier* ‘officer’; <ai**rr**in hauz> *Irrenhaus* ‘house for the mentally disabled’; <ʾe**ss**n> *essen* ‘eat’.

German <ck> is rendered in Ottoman-Turkish as <qk> in inlaut between vowels, probably thus imitating the two-grapheme cluster of the German original:

> <ārtišu**qk**e> *Artischocke* ‘artichoque’, <(fr)šilu**qk**in> *(ver)schlucken* ‘gobble’, <fir ʾe**qk**iġ> *viereckig* ‘quadrangular’.
>
> In morpheme/word-final position it is spelled as <-q>, despite German <ck>: <di**q** darmy> *Dickdarm* ‘colon’.

German written geminates in auslaut or syllable coda in general are not rendered as geminates:

> <ʾe**l**buġin> *Ellbogen* ‘elbow’, <mi**t**wuḫ> *Mittwoch* ‘Wednesday’; <fe**t**> *fett* ‘fat’; <b**t**> *Bett* ‘bed’.

The complete list of the correspondences of graphemes and phonemes/phones in the glossary would by far exceed the space allocated for this article: But a feature telling us something about the missing diligence in compiling the dictionary can be observed in the fact that sometimes the author slavishly clings to the German spelling: typos in

German words are faithfully rendered in the transcription. So e.g., when there is written *Pfirsche* instead of *Pfirsiche* in German, then we also find <pfirše> in the transcription, etc. (cf. Table 2-4 and Fig. 2-6):

p. 18: <pfirše> *Pfirsche* instead of *Pfirsiche* 'peaches'
p. 18: <hazeluns> *Haselunß* instead of *Haselnuß* 'hazelnut'
p. 18: <walluns> *Wallunß* instead of *Wallnuß* 'walnut'
p. 19: <munče> *Münze* 'coin' instead of *Minze* 'mint'
p. 33: <bešmussin> *beschmüßen* instead of *beschmutzen* 'to make dirty'
p. 35: <blissin> *blissen* instead of *blitzen* 'to flash (about a lightning)'
p. 39: <špečil ārč> *Speziel arzt* instead of *Spezialarzt* 'special medic'
p. 40: <geburst hlfr> *Gebürst helfer* instead of *Geburtshelfer* 'helper at giving birth'

Table 2-4: Misspellings in the 1916 dictionary.

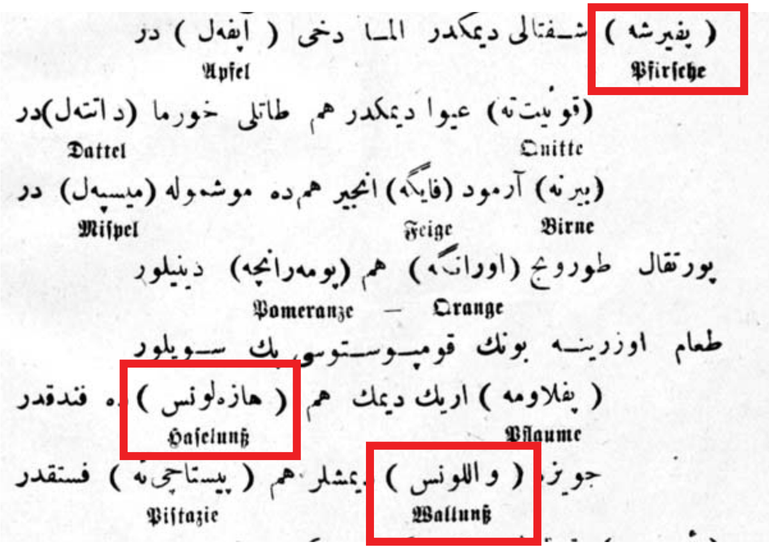


Fig. 2-6: *Almanca tuhfe*, p. 18.

The glossary gives the impression of a rather quickly written book. One intention will undoubtedly have been fame—as it was the first versified Ottoman-Turkish-German dictionary. Maybe Şerefeddin Pasha also wanted to exploit the Ottoman-German alliance in World War I to make some money.

At any rate, as a result we have to say that Şerefeddin Pasha did not succeed in making up a coherent system—even less an ideal system with a 1:1-relation between phoneme and grapheme. This glossary and another attempt at writing German with Arabic letters some 325 years earlier[36] have one characteristic feature in common: they seem to be both without predecessors and without followers. They are neither part of a tradition nor the beginning of one. And interestingly they both used different ways of transcribing German in their texts, which can be seen from a comparison of the spelling of identical words in both texts (Table 2-5).

[36] The anonymous manuscript written ca. 1590 (Flügel 2006 [A. F. 437] is kept in the Austrian National Library [*Österreichische Nationalbibliothek*], Vienna). The manuscript, which was stored in the National Library of Vienna already before 1610, consists of 162 folia (11 of them empty) in post-card format (17.1 x 11.2 cm), mostly inscribed on both sides with 15 lines, written on European (i.e. non-Turkish) paper. It contains texts in Arabic, Persian, Ottoman-Turkish, Bosnian (Serbo-Croatian), Latin (called 'Italian'), Hungarian and German, all written by the same hand in Arabic letters and presumably recording an oral presentation. So this is perhaps a final copy of a collection of texts showing quite a wide range of interests on the side of the anonymous collector (For more information about the possible cultural and ethnic background of the collector and/or writer see Römer 2014. Nevertheless, we have to admit that we do not know anything for sure. A hint might be the fact that the German Lutheran songs are not provided with any translation, while the Ten Commandments and the Lord's Prayer are. Some other [though smaller] European texts also remained untranslated.). The texts themselves are quite diverse: there are suras from the Qur'ān in Arabic, magic spells in Turkic, poetry in Persian, and other Oriental miscellanea. More on this see in Ivušić 2012, Bichlmeier/Ivušić 2013, Ivušić 2013, Mittwoch/Mordtmann 1927. For information about the possible place where the manuscript was written (from a Hungarologist's point of view), see Gragger 1927a, 1927b. For a first edition of the texts of the manuscript, see Mittwoch 1927. The manuscript is what might be called a *majmua* (Turk. *mecmua*; i.e. composite manuscript) in Orientalist terminology, but as such it is also a solitary unit—until now, no other manuscript showing such a variety of texts and languages has been found. But even more surprising are the texts in the above-mentioned European languages: the Ten Commandments and the Lord's Prayer in Latin, Hungarian, Bosnian and German; a love-poem in Bosnian (Serbo-Croatian); a few Hungarian songs and 19 German songs (most of them Lutheran church songs, but also a few secular ones); at least one of them has come down to us only in this manuscript. The German texts comprise clearly under 10 percent of the whole manuscript.

German lexeme	16th-century manuscript	Şerefeddin Pasha 1916
bist ‘[thou/you] art/are’	<p$_{i}$ist> /pist/ (Bavarian)	<bist>
Brot ‘bread’	<pr^{u}ut> /prōt/ (Bavarian)	<brut>
drei ‘three’	<dra^{a}i>	<dray>
du ‘thou, you (nom. sg.)’	<du^{u}>	<du>
eine(m) ‘one, a’	<āin^{a}m>	<āine>
euch ‘you (dat./acc. pl.)’	<auiḫ>, <a^{u}uiḫ>	<auiḫ>
Gottes ‘God’s’	<ġu^{u}tt^{a}s>	<ġutts>, <ġutas>
haben ‘have’	<ḥa^{a}b^{a}n>	<habin>, <habn>
Herr(en) ‘man (men), lord(s)’	<h^{a}r$_{i}$in>	<hr>, <her>
Herz(en) ‘heart(s)’	<h^{a}rč^{a}n>, <ḥ^{a}rč^{a}n>	<hrč>
Mutter ‘mother’	<mu^{ɔu}t^{a}r>	<mutter>
Namen ‘name(s)’	<na^{a}m^{a}n>	<namn>
und ‘and’	<au^{u}nd>	<^{c}und>
unser ‘our’	<au^{ɔu}nz^{a}r>	<aunzr>
Vater ‘father’	<fa^{a}t^{a}r>	<fater>, <fāter>

Table 2-5: Comparison of spellings of identical words in the sixteenth-century manuscript and in the dictionary of 1916.

6. Conclusions

Versified glossaries have a significant position in lexicology, educational sciences, and children’s literature. Therefore, such dictionaries deserve to be studied. Unlike prose dictionaries, versified dictionaries are not books of reference. Şerefeddin Pasha tried to avoid this problem by preparing a categorized dictionary, but he missed one of the main requirements of this tradition, i.e., the meter. This is one of the main problems of his book.

As we see in the French glossary and in other glossaries that appeared in the nineteenth century, parallel to the change in the reason for their production, the background of the poets, as teachers, also changed. The poet of the French versified glossary is a translator while the author of the German versified glossary is a medical doctor.

Finally, if we need to compare this specific glossary with the others that were written within this tradition, it could be claimed that Şerefeddin Pasha's work was not well prepared. And one of the reasons could be that he learned the German language too late and he was not very good at it.

In the end, we see that Şerefeddin Pasha's glossary is rather singular in at least two ways: as the first and only versified Ottoman-Turkish-German glossary and as a text containing one of the very few attempts at spelling German words with Arabic letters.

As it came simply too late, the 'German gift' was not able to create a tradition. After the founding of the Turkish Republic, the use of Arabic script for writing Turkish was abolished in 1928 in favor of the Latin script—thus there was no need any longer for taking on the trouble of using different alphabets. And, in accordance with Atatürk's state doctrine of westernizing Turkish society, the writing of this kind of traditional glossaries was not *en vogue* anymore; the European way of writing books for learning foreign languages and of writing dictionaries took hold step by step, ousting traditional concepts of doing so.

Being the first and only versified Ottoman-Turkish-German glossary, Şerefeddin Pasha's glossary is at the same time part of the end-game of this tradition of glossaries (the last one being a 1924 Turkish-Persian glossary)—thus deserving certainly more intensive research in the very near future.

Bibliography

Aksoy, Ömer Asım. 1959. "Şeyh Ahmed ve Nazmü'l-Leâl." *Türk Dili Araştırmaları Yıllığı Belleten*, 205-48.

Alkan, M. Nail. 2015. "Hayranlık, Dostluk ve Çıkar Üçgeninde Türk-Alman İlişkileri." *SDU Faculty of Arts and Sciences Journal of Social Sciences*, 34: 35-48.

Arslan, Anıl. 2016. *Tuhfe-i Hüsâmî: İnceleme, Çeviri Yazılı Metin, Dizin*. M.A. Thesis, Eskişehir Osmangazi University, Institute of Social Sciences.

Arslan, Mustafa. 2011. *Menâzımü'l-Cevâhir: Arapça-Farsça-Türkçe Manzum Sözlük*. Ankara: Genesis Kitap.

Bichlmeier, Harald, and Ivušić, Branka. 2013. "Zur dialektologischen Einordnung der deutschen Texte einer osmanischen Sammelhandschrift vom Ende des 16. Jh.s." In *Strömungen in der Entwicklung der Dialekte und ihrer Erforschung. Beiträge zur 11. Bayerisch-Österreichischen Dialektologentagung in Passau, September 2010* ed.

by Rüdiger Harnisch, Sigrid Graßl and Rosemarie Spannbauer-Pollmann, 365-84. Regensburg: edition vulpes.

Blau, Otto. 1868. *Bosnisch-türkische Sprachdenkmaler/gesammelt, gesichtet und herausgegeben von Otto Blau*. Leipzig: F.A. Brockhaus.

Çetin, Nihad M. 1991. "Aruz." In *Türkiye Diyanet Vakfı İslâm Ansiklopedisi*, vol. 3: 424-37. Istanbul: Türkiye Diyanet Vakfı.

Dankoff, Robert; Kut, A. Turgut; and Weitenberg, Joseph J. S. 1996. *The Versified Armenian-Turkish Glossary by Kalayi ca. 1800*. Cleveland: Cleveland State University Armenian Publications.

Doğan Averbek, Güler. 2018a. "Dillerinden Biri Türkçe Olan Manzum Sözlükler Üzerine Yapılan Çalışmalar Bibliyografyası." *Divan Edebiyatı Araştırmaları Dergisi*, 21: 85-114.

Doğan Averbek, Güler. 2018b. "Türkçe-Arnavutça Manzum Sözlük Dürre-i Manzûme'nin Bilinmeyen İki Nüshası." *FSM İlmî Araştırmalar İnsan ve Toplum Bilimleri Dergisi*, 12: 223-42.

Doğan Averbek, Güler. 2019. "Anadolu Sahasında Müstakil Bir Tür Olarak Manzum Sözlükler (Tuhfeler)." *Divan Edebiyatı Araştırmaları Dergisi*, 23: 62-83.

Doğan Averbek, Güler, and Bichlmeier, Harald. 2020. *Almanca Tuhfe—Deutsches Geschenk (1916). The only versified Turkish-German Dictionary with an Introduction on versified dictionaries as coursebooks*. Frankfurt/Wien/Bern etc.: Peter Lang.

Durmuş, İsmail. 2006. "Mukaddime." In *Türkiye Diyanet Vakfı İslâm Ansiklopedisi*, vol. 31: 115-17. Istanbul: Türkiye Diyanet Vakfı.

Elmalı, Hüseyin. 2001. "Kaside." In *Türkiye Diyanet Vakfı İslâm Ansiklopedisi*, vol. 24: 562-64. Istanbul: Türkiye Diyanet Vakfı.

Erik, Bayram Ali. 1982. *Tuhfe-i Nuriyye ve Zeyl-i Tuhfe-i Nuriyye*. Graduation Thesis, Hacettepe University.

Fedayî Mehmed Dede. 2019. *Tuhfe-i Fedayî*, ed. by Güler Doğan Averbek. Istanbul: Kitabevi Yayınları.

Freimark, P. 1960-2002. "Muḳaddima." In *Encyclopaedia of Islam, Second Edition*, vol. 7: 495-96. Leiden: Brill.

Gibb, E. J. W. 1900-09. *A History of Ottoman Poetry*. London: Luzac & Co.

Gragger, Robert. 1927a. "Der magyarische Text von Murād's ʿGlaubenshymnus' mit deutscher Übersetzung." In *Literaturdenkmäler aus Ungarns Türkenzeit. Nach Handschriften in Oxford und Wien*, edited by Franz Babinger, et al., 55-69. Berlin/Leipzig: Walter de Gruyter.

Gragger, Robert. 1927b. "Türkisch-ungarische Kulturbeziehungen". In *Literaturdenkmäler aus Ungarns Türkenzeit. Nach Handschriften in*

Oxford und Wien, edited by Franz Babinger, et al., 1-32. Berlin/Leipzig: Walter de Gruyter.

Güleç, İsmail, and Doğan Averbek, Güler. 2019. “Türk Çocuklar İçin Yazılmış Fransızca Manzum Bir Sözlük: Miftâh-ı Lisân.” *FSM İlmî Araştırmalar İnsan ve Toplum Bilimleri Dergisi*, 14: 53-80.

Haber. 19 December 1926. “İrtihal news.”

Hadžijahić, Muhamed. 1941. “Aljamiado Literatura (kod Hrvata).” In *Hrvatska Enciklopedija*. Encyclopaedia Croatica. Vol. I. *A-Automobil*, ed. by Mate Ujević, 300-01. Zagreb: Konzorcija Hrvatske enciklopedije.

Hadžijahić, Muhamed. 1955. “Književnost na arabici (aljamiado).” In *Enciklopedija Jugoslavije*. Vol. I: *A-Bosk*, ed. by Anto Babić, Miroslav Krleža, et al., 144-45. Zagreb: Izd. Nakl. Leksikografskoga Zavoda FNRJ.

Ivušić, Branka. 2012. “Die südslavischen Aljamiado-Texte der Wiener Sammelhandschrift Flügel 2006.” *Welt der Slaven*, 57/2: 380-98.

Ivušić, Branka. 2013. “*Sieben auf einen Streich.* Manuskript des Monats 04/2013.” Accessed on June 20 2019. https://www.manuscript-cultures.uni-hamburg.de/mom/2013_04_mom.html.

Kaya, Hasan and Ayçiçeği, Bünyamin. 2019. *Müsellesnâme: Osman Şâkir’in Manzum Sözlüğü*. Istanbul: İdeal Kültür Yayıncılık.

Kılıç, Atabey. 2007. “Türkçe-Arapça Manzum Sözlüklerden Sübha-i Sıbyân, 2: Metin.” *Turkish Studies*, 2 (1): 29-71.

Kontzi, Reinhold. 1980. “Aljamiado-Literatur.” In *Lexikon des Mittelalters I*: *Aachen bis Bettelsordenskirchen*, 415. Stuttgart: J.B. Metzler’sche Verlagsbuchhandlung.

Mittwoch, Eugen, and Mordtmann, Johannes Heinrich. 1927. “Die Wiener Sammelhandschrift.” In *Literaturdenkmäler aus Ungarns Türkenzeit. Nach Handschriften in Oxford und Wien*, edited by Franz Babinger, et al., 70-87. Berlin/Leipzig: Walter de Gruyter.

Mittwoch, Eugen. 1927. “Die deutschen, magyarischen, kroatischen und lateinischen Texte der Wiener Sammelhandschrift.” In *Literaturdenkmäler aus Ungarns Türkenzeit. Nach Handschriften in Oxford und Wien*, ed. by Franz Babinger, et al., 88-130. Berlin/Leipzig: Walter de Gruyter.

Muhtar, Cemal. 1993. *İki Kur’an Sözlüğü: Luğat-ı Ferişteoğlu ve Luğat-ı Kânûn-i İlâhî*, Istanbul: Marmara Üniversitesi İlahiyat Fakültesi Vakfı.

Neweklowsky, Gerhard. 1996. *Die bosnisch-herzegowinischen Muslime – Geschichte, Bräuche, Alltagskultur*. Klagenfurt/Salzburg: Wieser.

Ölker, Gökhan. 2004. “Tuhfetü’l-Uşşâk: Metin, Sözlük.” M.A. Paper, Selçuk University, Institute of Social Sciences.

Ölker, Gökhan. 2005. "Şiir Diliyle Rumca Öğrenimi: Tuhfetü'l-Uşşâk." *Gazi Üniversitesi Türkiyat Araştırmaları Merkezi I. Türkiyat Araştırmaları Sempozyumu,* (11-13 May 2005, Ankara).

Ölker, Gökhan. 2009. "Rumca-Türkçe Manzum Sözlük Tuhfetü'l-Uşşāk." *Turkish Studies*, 4 (4): 856-72.

Özkan, Ferudun Hakan. 2013. "Bozoklu Osman Şâkir ve Şehdü Elfâz'ı." *Uluslararası Sosyal Araştırmalar Dergisi,* 6 (26): 428-62.

Reyhan, Cenk. 2005. "Türk-Alman İlişkilerinin Tarihsel Arka Planı (1878-1914)." *Belleten,* LXIX (254): 1-46.

Römer, Claudia. 2014. "Cultural Assimilation of a 16th-Century New Muslim. The Mecmua ÖNB A. F. 437." In *CIÉPO 19. Osmanlı Öncesi ve Dönemi Tarihi Araştırmaları*. 2 vols., edited by İlhan Şahin, Hikari Egawa, Emine Erdoğan Özünlü, Tuncay Öğün, 607-019. Istanbul: İstanbul Esnaf ve Sanatkarlar Odaları Birliği Yayını.

Rossi, Ettore. 1946. "Notizia Su Un Manoscritto Del Canzoniere Di Neẓīm (Secolo xvii-xviii) In Caratteri Arabi e Lingua Albanese." *Rivista degli studi orientali*, 21/2-4 (November): 219-46.

Şerefeddin Pasha. 1916. *Almanca tuḥfe* [Ālmanje tḥfe]. *Deutsches Geschenk*. Istanbul: Matbaa-i Osmaniye.

Siebs, Theodor. 1905. *Deutsche Bühnenaussprache*. Dritte Aufl. Berlin/Köln/Leipzig: Albert Ahn.

Siebs, Theodor. 1915. *Deutsche Bühnenaussprache*. Elfte Aufl. Bonn: Albert Ahn.

State Archives of Turkish Republic, 253-57, Y.PRK.ASK.

Turan, Muhittin. 2012. "Hasan Rızâyî ve Kân-ı Ma'ânî İsimli Manzum Sözlüğü." *Turkish Studies,* 7 (4): 2939-992.

Uskufi, Muhamed Hevai. 2001. *Makbūl-i 'Arif: Potur Šahidiya*. Tuzla: Općina Tuzla Behram-begova biblioteka, Narodna i univerzitetska biblioteka.

Vehbî, Sülbülzâde. 2012. *Tuhfe-i Vehbî: Metin, Dizin, Tıpkıbasım*. Ed. Ahmet Yenikale, Kahramanmaraş: Öncü Basımevi.

Verburg, Antoinette C. 1997. "The Tuhfe-i Şāhidī: A Sixteenth Century Persian-Ottoman Dictionary in Rhyme." *Archivum Ottomanicum*, XV: 5-87.

Yıldırım, Nuran. 2006. "Sağlık Kuruluşları." In *Surların Öte Yanı: Zeytinburnu*, 3rd ed., 296-355. Istanbul: Zeytinburnu Kültür Yayınları.

Yurtseven, Necmettin. 2003. *Türk Edebiyatında Arapça-Türkçe Manzum Lügatlar ve Sünbülzade Vehbi'nin Nuhbe'si*. M.A. Thesis, Ankara University, Institute of Social Sciences.

Yûsuf Hâlis. 2006. *Miftâh-ı Lisân: Manzum Fransızca-Türkçe Sözlük*, ed. Mehmet Kırbıyık, Istanbul: Beşir Yayınevi.

Chapter Three

Old Frisian Terms in the *Deutsches Rechtswörterbuch*. The Multilingual Approach of the Dictionary of Historical German Legal Terms[1]

Andreas Deutsch

Abstract

The *Deutsches Rechtswörterbuch* (translated literally: German Legal Dictionary) is the most comprehensive dictionary of German technical terminology. Despite its name, it focuses exclusively on older legal language. Moreover, not only does the dictionary contain German historical legal terms, it also includes legal terms from all West Germanic languages. It covers the legally relevant terms from the beginning of Germanic written tradition (about 400 AD) up to 1815 (word occurrences in the online version until 1835). From the perspective of legal historical research, Old Frisian is of special interest. This language has been handed down mainly in numerous legal sources, which contain a specific vocabulary and provide an exciting insight into past living conditions—demonstrated by words such as "sinekerf" (Sehnekerb, punishable dissection of a tendon), "sawenbethe" (Siebenbuße, sevenfold fine) or "sponste" (Sponst, seduction). Once completed, the *Deutsches Rechtswörterbuch* is expected to consist of sixteen volumes with about 120,000 main entries. About 100,000 main entries in more than 20,000 columns have been printed so far—in alphabetical order from Aachenfahrt (pilgrimage to the coronation church of the Holy Roman Empire at Aachen) to Stegrecht (tax on goods which are

[1] The article is based on a lecture delivered by the author at the ICHLL in Leeuwarden, 13 June 2019. The style of the oral speech has largely been retained.

unloaded from or loaded onto ships). Every year more than a thousand new entries will be added.

1. Introduction

"It is not difficult to believe that law and poetry rose from the same bed,"[2] is how philologist and legal historian Jacob Grimm emphasized the common roots of law and literature in his essay "Von der Poesie im Recht" ("On Poetry in Law", Grimm 1816, 27). The elder of the two Brothers Grimm, who together collected the fairy tales and started the *Deutsches Wörterbuch* ("German Dictionary") in 1838, was one of the visionaries of that nineteenth-century scholarly movement that ventured to explore the history and law of the Germanic tribes. The group believed that here the origins of their own legal history could be found.[3] Since hardly any sources from the earliest Germanic period had been handed down, the researchers resorted to a number of medieval sources, which they considered to be the essence of natural language (cf., e.g. Eichhorn 1834, 283-284). In his *Deutsche Rechtsalterthümer* ("German Legal Antiquities"), first published in 1828, Jacob Grimm assembled early legal sources and revealed linguistic connections on the basis of exemplary legal terms (Grimm 1899).

Thus it was the "German" language that was to serve as glue. With the term "German" the Brothers Grimm (like many other nineteenth-century scholars) denoted the entire West Germanic language family, as Jacob Grimm had defined it in his introduction to the German Dictionary: "from the German language...the ancient Gothic tribe is excluded, as is the Nordic or Scandinavian one, so that the Frisian, Dutch, Old Saxon and Anglo-Saxon are still included in the German language in a narrow sense".[4]

Based on Grimm's assumptions, but also on those of Karl Friedrich Eichhorn, Wilhelm Eduard Wilda and other authoritative scholars, the Munich-based legal historian Karl von Amira developed the concept of a systematic analysis of what he considered to be Common Germanic law. In his "Grundriss des germanischen Rechts" ("Outline of Germanic Law"), published in 1893, he was not concerned with similarities in content of individual legal rules as evidence of a legal relationship. Rather, according

[2] "Dasz recht und poesie miteinander aus einem bette aufgestanden waren, hält nicht schwer zu glauben".
[3] On the problems regarding this hypothesis cf. Kroeschell 1995, 65.
[4] "man scheidet von der deutschen sprache…den alten gothischen stamm aus, als den nordischen oder scandinavischen, so dasz gleichwohl die friesische, niederländische, altsächsische und angelsächsische noch der deutschen sprache in engerm sinn zufallen", *Deutsches Wörterbuch* (DWB), vol. I, Leipzig 1854, p. XIV.

to Amira, "the only, albeit merely comparatively reliable standard in the assumption is that the legal families of previous times coincide with the language families (East and West Germanic, Gothic in a broader sense and Scandinavian...etc.). The language families are the expression of a historical relationship among peoples".[5] This special relationship between law and language was one of the reasons for founding the *Deutsches Rechtswörterbuch* in 1897.

From the beginning, Old Frisian law played a special role for the researchers of the so-called Germanic law (Deutsch 2012, 26). In 1828 Jacob Grimm explained in his "Rechtsalterthümer" that the most significant indication of a text's particularly old age and of its true primitive nature was the imagery and poetry which it contained (Grimm 1899, X). According to Grimm, when comparing two similar texts, the more poetic one is always the older text. Regarding Old Frisian law Grimm (1816, 36) said: "The law book of the old Frisians contains throughout its oldest and best part pure and precious poetry."[6] Therefore, Grimm considered it even older than the Anglo-Saxon laws.

After the East Frisian public official Tilemann Dothias Wiarda had edited certain sources of Frisian law (Wiarda 1805, Wiarda 1820), Karl Freiherr von Richthofen, a close pupil of Grimm, was the first to publish a collection of Old Frisian legal sources in 1840 (Richthofen 1840a).[7] As can be seen in the preface to his "Old Frisian Dictionary", Richthofen, too, regarded the Frisian texts primarily as a "rich repository of the older German law" (Richthofen 1840b, Introduction).[8] He had begun to investigate them because "the study of Frisian folk life is particularly important for the understanding of German history" since the "legal monuments of Frisia...are

[5] "der einzige, wenn auch nur relativ verlässige Maßstab in dem Satze gegeben, daß die Rechtsfamilien der älteren Zeit sich mit den Sprachfamilien (ost- und westgermanisch, gotisch i.w.S. und skandinavisch…usf.) decken. Die Sprachfamilien sind der Ausdruck der geschichtlichen Verwandtschaft unter den Völkern"—cf. Amira 1913, 8. The first edition appeared in 1893 under the title "Recht".

[6] "Der alten Friesen Gesetzbuch hat durchgehends seinem altem besten Stück nach, poetisches Schrot und Korn in sich." On the specific style of the Old Frisian laws cf. Bremmer 2011.

[7] Of course, there were some other forerunners: Matthias von Wicht (1746) had edited the *Ostfriesisches Landrecht* (1515/20), which essentially contains Frisian law, but linguistically belongs to Low German (the manuscript used by Wicht dates from 1562).

[8] "dass ohne ein solches [wörterbuch] jene reichen fundgruben des älteren deutschen rechtes, deren reiche flötze noch völlig unausgebeutet liegen, nur wenig zugänglich bleiben würden".

available in such abundance as in no other German tribe.” [9] Deeply impressed, he compared his work to the exploitation of a mine, the seams of which are still utterly unexploited.

2. The *Deutsches Rechtswörterbuch*

2.1 Some facts

The *Deutsches Rechtswörterbuch* (*DRW*, translated literally: German Legal Dictionary) is the most comprehensive dictionary of German technical terminology. Despite its name, it focuses exclusively on older legal language. Moreover, not only does the dictionary contain German historical legal terms, it also includes legal terms from all the other West Germanic languages. It covers the legally relevant terms from the beginning of the Germanic written tradition up to the nineteenth century.

Almost 100,000 main entries have been printed so far—in alphabetical order from “Aachenfahrt” (pilgrimage to the coronation church of the Holy Roman Empire at Aachen) to “Stegrecht” (tax on goods which are unloaded from or loaded onto ships). The dictionary currently covers about 500,000 quotations and short references in 20,000 print columns. Every year more than a thousand new main entries are being added. Once completed, the *DRW* is expected to consist of sixteen volumes with about 120,000 main entries (Deutsch 2016, Deutsch 2011; Kirkness 2016, 77).

In addition to the lemma and the indication of the part of speech as well as further linguistic details (if necessary), each main entry contains an explanation of the legally relevant meanings of the term and corresponding citations of the sources, providing references and evidence of the word's existence and use. These quotations cover the temporal and geographical occurrence of the word and of its respective meaning within the limited space available in the dictionary.

The basis for the creation of the dictionary are about two and a half million quotation slips, which researchers collected starting around 1900, with about 8500 citation keys—referring to entire collections of texts or single sources. The electronic text archive with about 30 million words is still growing. It is supplemented by 5.4 million word forms from existing quotations of the *DRW*.

---

[9] “das studium des friesischen Volkslebens [<sei>] für das verstehen deutscher geschichte im umfassendsten sinne dieses wortes, wichtig, die „rechtsdenkmäler Frieslands…sind in solchem reichthum vorhanden, wie bei keinem andern deutschen stamme”.

Regarding the content of the dictionary, the *Deutsches Rechtswörterbuch* obviously covers primarily the technical language of the law. However, the *DRW* pays equal attention to the legally relevant parts of everyday language (Künßberg 1932, IX; Deutsch 2010, 423). This includes, for example, all coins, measures, weights, taxes and charges. Likewise, all professions are being covered insofar as they are connected with official or other legal functions or organized in a guild or a similar association. It seems that the latter is of interest in particular in onomastics, because many professions from the past have been preserved in surnames (Deutsch 2017). Furthermore, both the broad special vocabulary of commerce and trade, namely that of the Hanseatic League, and the vocabulary of mining rights are dealt with in the *DRW*.

The time span covered by the *Deutsches Rechtswörterbuch* is extremely broad compared to other dictionaries: About 1400 years are considered—from the beginnings of written documentation up to a fixed time limit, which has been slightly altered several times. The earliest records at the moment are a number of words culled from Wulfila's Gothic Bible translation, dating from 384 AD, which is used as a comparative source. Two records for the entries "Graf" and "Mundburt" originate from a Merovingian charter of Clovis I, dating from the year 479 AD. Apart from this, Old English texts are among the earliest sources. In order for a compound to be added to the dictionary, its first verifiable attestation must date from the year 1700 or earlier, while a simplex is included when it is attested not later than 1815. If the first record of a term does not fall within the aforementioned time limits but still dates from before 1835, a short reference to the term will be made, which does not get printed but is published in the freely accessible online version (www.deutsches-rechtswoerterbuch.de; cf. Bedenbender 2014; Kirkness 2016, 78; Storrer 2014, 1246).

The linguistic framework of the project is the "German language", to which—according to the concept of the nineteenth century—the whole of the West Germanic language family belongs, as Jacob Grimm laid down in his introduction to the "German dictionary". This framework includes vernacular words found in Latin texts from among the Germanic tribes of the migration period, the so-called "Leges barbarorum" (450-800). It encompasses the following languages and periods: Old English (500-1100), Old High German (600-1050), Lombardic (650-1000), Old Dutch (700-1200), Old Saxon (800-1200), Old Frisian (800-1550), Middle High German (1050-1350), Middle Dutch (1200-1600), Middle Low German (1200-1650), Early Modern German (1350-1650) and Modern High

German (from 1650 onwards) (Falkson, Lemberg and Lill 2002; Lemberg 2007).

To comprise so many languages and language periods in one dictionary poses a special challenge. The main entries must not only cover as many time periods and regions as possible, but also all the languages in which the word appears within the material of the *Deutsches Rechtswörterbuch*.

2.2 The example "setzen" (to set, to place, to put, to seat)

Moreover, as there was no established orthography in the old days, this also means that a large number of spellings, which in some cases can hardly be overseen, must be taken into account. Thus, for the linguistically simple word "setzen" (set) there are more than a hundred spellings considered in the *Deutsches Rechtswörterbuch*. Incidentally, the meaning of the verb is self-explanatory only at first glance. In the *DRW* a total of thirty-nine word meanings were included, even though—of course—only the legally relevant ones were taken into account.[10]

gesaczet	gesetzit	ghesettet	sazte	settede	setzte
gesaczt	gesetzt	gúesezzet	sazten	setteden	setzten
gesaczten	gesetzte	gvisezzet	secczen	setten	setzzen
gesat	gesetzteme	sad	secz	settende	seydcin
gesatt	gesetzten	saß	seczcen	settene	sez
gesatz	gesetzter	saste	secze	settens	sezen
gesatzet	gesezet	sat	seczen	settet	sezent
gesatzt	gesezt	sateden	seczin	setton	sezin
gesazt	gesezzen	satta	seczn	setzcen	sezt
geseczet	gesezzet	satte	seczt	setze	sezzen
geseczt	geßaczt	satten	seczzent	setzen	sezzet
geset	gezesezt	satzend	setczen	setzende	sezzüt
gesettes	gezet	satzt	sethten	setzene	sötzen
gesettet	ghesad	satzte	sett	setzet	sözen
gesetton	ghesast	satzten	setta	setzin	sözt
gesetzen	ghesat	satztend	settan	setzn	zetczn
gesetzet	gheset	satztú	sette	setzt	zetten

[10] Art. "setzen", *DRW* XIII, 408-424; https://drw-www.adw.uni-heidelberg.de/drw-cgi/zeige?term=setzen.

In order to ensure a clear sorting and grouping of the words in the dictionary and to provide a reliable way to find the words despite the many spellings, all entries are always provided with a Modern High German entry. This applies even to words which do not exist in Modern High German. For them, an artificial lemma is formed which is put in parentheses. In addition, the online version enables a search for any spellings of a word, so that special language skills are not required in order to look up, for example, Old English or Old Frisian words.

2.3 The example "Schultheiß"

An even more impressive example is the noun "Schultheiß" with its short form "Schulze". Again, the *Deutsches Rechtswörterbuch* encompasses more than a hundred spellings from different regions of Europe.[11]

escultaizo	schulcze	schultheisen	schultzheisen	sculteiten
scelta	schuldacio	schultheiß	schulz	sculteizin
sceltata	schuldheißen	schultheissen	schulzen	sculten
schautetten	schuldtheyß	schultheißen	scolten	sculteten
schelta	schulteis	schultheisz	scout	sculteto
scheltata	schulteise	schultheize	scoutate	scultetus
scheltum	schulteiss	schultheizen	scoute	scultezi
scholt	schulteize	schultheizsen	scouten	scultezin
scholteissen	schulten	schultheizze	scouthate	sculthacio
scholteisz	schúlteß	schulthes	scudassius	sculthaizium
scholteti	schultessen	schulthessen	sculdahis	scultheite
scholtheisen	schultesz	schulthesz	sculdais	scultheiten
scholtis	schultete	schultheysen	sculdaissihis	scultheiten
scholtiss	schulteyze	schultheyss	sculdasius	scultheti
scholtisse	schultez	schultheyßen	sculdassi	scultheto
scholtissen	schultezen	schultheyze	sculdhais	scultichten
scholtus	schulthaissen	schultis	sculdhaisus	skelta
schoute	schulthaize	schultisse	sculdheyzen	skeltata
schoutet	schůlthaizen	schultys	sculta	solteto
schoyltesse	schulthaizzen	schultze	scultaiz	zulheysse
schoyltessen	schultheis	schultzen	scultechte	*(et al.)*

[11] Art. "Schultheiß" *DRW* XIII, 1416-1428; https://drw-www.adw.uni-heidelberg.de/drw-cgi/zeige?term=Schulthei%DF&.

Today a rather uncommon term for an official, variants of "Schultheiß" still occur as a wide-spread last name. For instance, three current German politicians bear this name.[12]

As mentioned at the beginning, one of the fundamental assumptions of the founders of the *Deutsches Rechtswörterbuch* was that a historical language area also corresponded with a historical legal area, so that a more or less uniform West Germanic law was to be found within the West Germanic language area. Among other things, in order to find evidence for this assumption, the *Deutsches Rechtswörterbuch* was once established. Following this hypothesis all the way through, every legal term in the entire West Germanic language area had approximately the same meaning.

The word "Schultheiß" serves as a good example to illustrate this thesis, because no matter whether in the Lombard region, in Switzerland or in the Baltics: the noun means basically the same everywhere. As a core meaning, the *Deutsches Rechtswörterbuch* defines "Schultheiß" as follows: "A public official vested with different competences at different times and places, who are usually entrusted with the jurisdiction over a specific group of people and the administration of a district."[13]

At the same time, however, the main entry "Schultheiß" clearly shows the limits of the thesis of the dictionary's founding fathers. For apart from this very general explanation, very different specific meanings have developed in various regions and languages. Thus, the *Deutsches Rechtswörterbuch* differentiates between seventeen different officials under the first meaning of the entry "Schultheiß", with derived meanings to be added.

In chronological order there is first of all the Langobardic "Schultheiß". He is a "royal official (locopositus, rector loci) with a lower jurisdiction and military function".[14] In Middle-Dutch Law, however, the "scout" is a "vice-

---

[12] Martin Schulz, former Leader of the German Social Democratic Party, former President of the European Parliament; Olaf Scholz, German Vice Chancellor and Federal Minister of Finance; Svenja Schulze, German Federal Minister of the Environment, Nature Conservation and Nuclear Safety. This finds its equivalent in the word "schout" in modern Dutch. This word is also wide spread as a surname in the Netherlands: the Dutch Minister of Agriculture, Nature and Food Quality, for example, is named Carola Schouten. The examples could be continued.

[13] "ein Amtsträger mit zeitlich und örtlich unterschiedlicher Kompetenz, der idR. für die Gerichtsbarkeit über einen best. Personenkreis und die Verwaltung eines Amtsbezirks zuständig ist".

[14] "I 1. im langob. Recht: königliche Amtsperson (locopositus, rector loci) mit niederer Gerichtsbarkeit und militärischer Funktion in einem Unterbezirk eines Herzogtums (Schultheißei (I)); auch Vollstreckungsbeamter."

châtelain; amongst other things, with judicial and executional functions".[15] Old Frisian provides yet another specific meaning: "(noble) deputy of a count exercising the judicial, administrative and military authority in a district extending over several villages; he has presidency and power of execution at his court and at the district assembly as a representative of the count; he also charges fines and taxes".[16]

"Setzen" and "Schultheiß" are examples of legal terms that are more or less attested throughout the entire West Germanic language area and through all the centuries. However, and understandably so, most of the legal terms have been common only at certain times and in certain regions.

3. The Old Frisian legal terminology

For the following examples let us concentrate on Old Frisian legal terminology; not only as a tribute to the *genius loci*, but also because the Old Frisian sources are particularly interesting from the perspective of legal history (Algra 2011, 556; Popkema 2010; Köbler 2008). They do not just represent a specific vocabulary, but additionally provide an exciting insight into past living conditions. Tilemann Dothias Wiarda pointed this asset out as early as 1786 (Wiarda 1786, VIII-X). As already mentioned, it also aroused the curiosity of nineteenth-century scholars, starting with Grimm and Richthofen. Yet this aspect also becomes evident when looking at modern secondary literature.[17] Impeding a closer look at the Old Frisian language monuments is the lack of a comprehensive Old Frisian reference dictionary (Popkema 2012). Far too many Old Frisian words are still a mystery to us.[18] This large gap can neither be filled by the *DRW* nor by the detailed but very specific *Vollständiges Wörterbuch zum westerlauwersschen Jus Municipale Frisonum* ("comprehensive dictionary to the West Frisian Jus Municipale Frisonum"), compiled by Buma (1996) nor by the superb *Altfriesisches Handwörterbuch* ("Concise Old Frisian Dictionary") by

[15] "I 13 in der Kastellanei Brügge: Unterburgvogt; ua. mit gerichtlichen und Vollstreckungsaufgaben; auch als: vri scouttete. "

[16] "I 2. im afries. Recht: (adliger) Stellvertreter eines Grafen, der in einem über mehrere Bauerschaften reichenden Amtsbezirk die gerichtliche, administrative und militärische Gewalt ausübt; er hat Vorsitz und Vollstreckungsgewalt beim Schultheißengericht (I) und als Vertreter des Grafen beim Gauding, während der Asega als Urteiler fungiert; erhebt Bannbußen und Steuern."

[17] It would be presumptuous to resort to individual titles here.

[18] Some impressive examples are provided by Oebele Vries' papers, such as Vries 2017, Vries 2012 and Vries 1996.

Hofmann and Popkema (2008). Especially these last two dictionaries[19], however, potentially provide an excellent basis for compiling a comprehensive reference dictionary.

Looking into the Old Frisian sources, one may at times get the impression that the medieval Frisians only hurt and harmed each other. Presumably, however, they only had a particularly elaborate system of sanction and compensation (as already described by Wiarda 1786, VIII; cf. in detail Nijdam 2008; Nijdam 2014; Algra 2000, 259). In fact, this system resembles the catalogues of compositions tariffs[20] which have been handed down in several Germanic tribal law codes, drafted in Latin with several German inserts (the so-called *Leges Barbarorum*) as well as in the vernacular Anglo-Saxon laws (cf. Schumann 2012; Nehlsen 2008; Willoweit 2007): The perpetrator was not physically punished, but had to pay the aggrieved party a compensation for the injury caused. The amount was determined by the severity of the infringement, as described in minute detail in numerous Frisian legal rules.

Thus a special compensation was due for “Sehnekerb”, the punishable cutting through a tendon,[21] or for a bodily injury, which caused a “Speichelsprung” (saliva spring), which means a “wound, resulting in saliva flowing from the mouth”.[22] Special rules were also recorded for “Soltat” (*soldede*), a form of grievous bodily harm, which results in the victim soiling himself with his own excretions; a so-called “act of sullying”.[23] Having a concise meaning, this term can illustrate the significance of source citations in the *DRW*, for which purpose the source quotations of “Soltat” are rendered below:[24]

---

[19] In addition, cf. the Old Frisian Etymological Dictionary (Boutkan and Siebinga 2005) can prove to be very helpful for etymological questions, as well as the introduction to Old Frisian by Bremmer (2009). Mention should also be made of Faltings’ etymological dictionary of Frisian adjectives (Faltings 2010).

[20] Bußenkataloge, Kompositionensystem.

[21] Art. “Sehnekerb”, *DRW* XIII, 215: “afries.; Zerschneiden einer Sehne; als bußwürdige Tat”.

[22] Art. “Speichelsprung“, *DRW* XIII, 857: “afries.: Speichelfluss; Wunde, infolge deren Speichel aus dem Mund fließt”; cf. Nauta (1941) 112, 152.

[23] Art. “Soltat”, *DRW* XIII, 717: “im fries. Recht: schwere Körperverletzung, in deren Folge sich das Opfer mit den eigenen Ausscheidungen besudelt; sog. Sudeltat.”

[24] It would go beyond the scope of this lexicographical article to deal with the word itself and its etymology, therefore cf. Boutkan 2002; Nauta 1941, 151, Munske 1973, 142; His 1901, 324.

- *fonre* soldede*: hwersa en mon werth bifuchten and hi clagie ene* soldede*, and thi redieua thet rede, and hi sterwe binna thrim wikem, sa ielde ma hine mith ene fulle ielde; liwath hi ther vr, sa vndvnge ma mith tuelef ethum* [von der Sudeltat: wenn ein Mann angegriffen wird und er wegen einer Sudeltat Klage erhebt, und der Redjeve das bestätigt, und jener binnen drei Wochen stirbt, so büße man ihn mit einem vollen Wergelde; lebt er darüber hinaus, so entgehe man (der Klage) mit zwölf Eiden]*, Ende 13. Jh. BrokmerR. 50.*
- *thet is en* solded*, wersa mane mon alsa fir onfiucht, thet him sin spise eta tuam enden vngunge* [das ist eine Sudeltat, wenn man einen Mann so heftig angreift, dass ihm die Speise an den beiden Enden hinausgeht], *um 1300 HunsingoR. 52.*
- *huersa ma en* soldede *deth, aldermithe deth ma thene haghesta dadsuima* [wenn man eine Sudeltat verübt, erfüllt man damit (zugleich den Tatbestand der) tiefsten Bewusstlosigkeit], *14. Jh. EmsigerR. 70.*
- *ther ne meij ma nene* soldede *dua, the ne se thi hagesta dathswima den* [man kann keine Sudeltat begehen, wo nicht der Tatbestand der tiefsten Bewusstlosigkeit erfüllt wird], *Mitte 15. Jh. EmsigerR. 118.*
- soldede *schall men betheren myth viff marck vnd twe ensa. dith schall men whar maken myth einen veheeede. 1525 (Hs.) OstfriesRQ. 174.*
- soldede *schall man betheren myth viff marck vnd twe enßa. dith schall men whar maken myth einen vehe-eede. nach 1565 WurstfriesRüstrLR. 101.*
- soldede*, schwimsleck, bardes homlinck, knop vorbrandt, ein inretene wunde, sleitota ahn de vorderen handt, dath vornste lith ahn de fyngeren vnd tanen, vnd ein schwarte schwange: alle viff marck twe enza. 2. Hälfte 16. Jh. OstfriesRQ. 204.*

For the *DRW*, as for any citation-based descriptive dictionary, the source quotations (and their wider context) are the most important basis for determining the word meaning (Lemberg 1998). They also provide the reader with information on the temporal and regional spreading of the word in question. Regarding the citations of “Soltat” in the *DRW*, the word can be traced from the thirteenth to the sixteenth century and occurs in various parts of Friesland, but not elsewhere.

The elaborateness of the aforementioned Old Frisian sanction system is also reflected in special terms for various precisely defined “Bußen” (boten) imposed on the perpetrator. In order to restore the peace between the parties involved and to compensate the victim or their heirs appropriately and justly, a comprehensive gradation system was applied (Nehlsen 1983), as numerous entries of the *DRW* illustrate. Thus, in one of the most recent fascicles of the dictionary the entry "Siebenbuße" occurs, which means: "Sevenfold amount of a fine (as compensation for the victim) for serious offences, in particular for breach of the peace."[25]

In addition, the entries "Dreizehnbuße" (thirteenfold amount of a fine), "Fünfbuße" (fivefold amount of a fine), "Neunbuße" (ninefold amount of a fine) "and "Sechsbuße" (sixfold amount of a fine) are already printed.[26] The importance of medieval number symbolism (cf. Suntrup 1998) is clearly evident here. Apart from divisors in the duodecimal system, the prime numbers dominate here.

These numbers are also encountered in other areas of Old Frisian law. Let me single out just a few examples of the significance of the number seven: "Siebenteil" (one of seven shares of an amount to be paid as a compensation for manslaughter – as part of an expiation contract),[27] "Sieben Küren" (a term for the seven Superior Statutes, which list the rules of the Upstalsboom League)[28] and the "Siebenhardenbeliebung" (the North Frisian territorial law agreed upon and written down on the island of Föhr in 1426 by the representatives of the seven judicial and administrative districts of North Frisia).[29] This accumulation of the number seven is certainly no coincidence, as in the medieval world the number seven stands for the union of heaven and earth as a perfect and sanctified number. From the seven days of the week to the seven planets known in the Middle Ages to the seven liberal arts, the number also played a predominant role outside the law (Wetzel 2011, 304; Ernst 2007).

Attributing a special significance to certain numbers is a phenomenon that can be detected in legal sources of the entire West Germanic language area,[30] as the detailed *DRW* entries on numerals such as "five", "six",

---

[25] Art. "Siebenbuße", *DRW* XIII, 484: "Bußzahlung in siebenfacher Höhe für schwere Straftaten, insb. Bruch des Friedens".
[26] *DRW* II, 1105; *DRW* III, 1069; *DRW* IX, 1456; *DRW* XIII, 123.
[27] Art. "Siebenteil", *DRW* XIII, 495-496, II.: "einer von sieben Anteilen an einer Sühnezahlung für Totschlag."
[28] Art. "[2]sieben", *DRW* XIII, 473-483, XVIII: "im afries. Recht: sieben Küren Bez. für die sieben Überküren, die ua. die Verfassung des Bundes von Upstalsboom enthalthen".
[29] Art. "Siebenhardenbeliebung", *DRW* XIII, 489.
[30] cf. Grimm 1899, 285–310. Number symbolism can also be observed in legal iconography. Just think of the octagonal crown of the Holy Roman Empire with its

"seven", "nine", "eleven" or "seventeen" impressively show—as even numerals are dealt with in the dictionary when the numbers bear a special legal significance. This also applies to the article "dreizehn" (thirteen)—my last example. Just as there were thirteen sworn representatives of the citizenship in the town hall of Leeuwarden, we find thirteen councillors in Augsburg and thirteen jurors in the Swiss canton of Schwyz.[31] A closer comparison of these examples shows that the number thirteen was probably chosen for pragmatic reasons, since, unlike the usual twelve-member bodies, a majority is always guaranteed. This shows that the all-over view of the entire West Germanic legal and linguistic area enabled by the *DRW* can provide insights even beyond the actual meaning of words.

4. Conclusion

When the "Dictionary of Historical German Legal Terms" (*DRW*) was founded in 1897, one of its purposes was to uncover the legal-historical similarities and connections in the West Germanic language area. It therefore not only deals with the legally relevant vocabulary of German, but also takes into account all West Germanic languages, including Old Frisian. Meanwhile the fourteenth volume of the dictionary is in progress. Around 100,000 main entries have been printed (from A to the end of S). The dictionary is used daily as a reference work in various countries around the world, not least via the freely accessible online version (www.deutsches-rechtswoerterbuch.de). In the case of some legal words, it can indeed be shown that they were used in a comparable manner throughout the West Germanic linguistic area (e.g. "Schultheiß"). However, many legal words explained in the legal dictionary only occur in a single, narrow legal and linguistic area, as can be demonstrated by the Old Frisian legal vocabulary in particular.

complex arrangement of the total of 120 gems and 144 large and 96 small pearls. The four plates of the octagonal imperial crown made of gold plate are framed with 14 pearls each (twice seven). There are 26 beads each on the two side plates (two times 13), and on the forehead and neck plate there are 18 beads (two times nine), which add up to 144 (twelve times twelve), the number that refers to the celestial Jerusalem; cf. Decker-Hauff 1955, 589-590.

31 Art. "dreizehn", *DRW* II, 1105.

Bibliography

Algra, Nikolaas Egbert. 2000. *Oudfries Recht 800-1256*. Leeuwarden: Fryske Akademy.

Algra, Nikolaas Egbert. 2011. “Grundzüge des friesischen Rechts im Mittelalter”. In *Handbuch des Friesischen/Handbook of Frisian Studies*, edited by Horst Haider Munske, 555-570. Tübingen: Niemeyer.

Amira, Karl von. 1893. “Recht.” In *Grundriss der germanischen Philologie*, edited by Hermann Paul, vol. III, Straßburg: Trübner.

Amira, Karl von. 1913. *Grundriss des germanischen Rechts*, 3rd edition. Straßburg: Trübner.

Bedenbender, Almuth. 2014. “Das Deutsche Rechtswörterbuch im Netz.” In *Vernetzungsstrategien, Zugriffsstrukturen und automatisch ermittelte Angaben in Internetwörterbüchern*, edited by Andrea Abel and Lothar Lemnitzer. Mannheim: IDS, online: https://pub.ids-mannheim.de//laufend/opal/pdf/opal2014-2.pdf#page=22.

Boutkan, Dirk. 2002. “Some Considerations on the Etymology of Old Frisian Sol-dede ‘Sudeltat’.” In *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis*, 7/1: 39-49.

Boutkan, Dirk and Sjoerd Michiel Siebinga. 2005. *Old Frisian Etymological Dictionary*. Leiden/Boston: Brill.

Bremmer Jr, Rolf H. 2009. *An Introduction to Old Frisian: History, Grammar, Reader, Glossary*, Amsterdam/Philadelphia: John Benjamins.

Bremmer Jr, Rolf H. 2011. “Dealing Dooms: Alliteration in the Old Frisian Laws.” In *Alliteration in Culture*, edited by Jonathan Roper, 74-92. Basingstoke/New York: Palgrave Macmillan.

Buma, Wybren Jan. 1996. *Vollständiges Wörterbuch zum westerlauwersschen Jus Municipale Frisonum*, Leeuwarden: Fryske Akademy.

Decker-Hauff, Hansmartin. 1955. “Die Reichskrone.” In *Herrschaftszeichen und Staatssymbolik*, edited by Percy Ernst Schramm, vol. II, 560-637. Stuttgart: Hiersemann.

Deutsch, Andreas. 2010. “Das Deutsche Rechtswörterbuch – ein Sprachwörterbuch zur Rechts- und Kulturgeschichte.” *Mitteilungen des Deutschen Germanistenverbandes,* 57: 419-431.

Deutsch, Andreas. 2011. “The ‘Dictionary of Historical German Legal Terms and its European Concept’.” In *The Fifth International Conference on Historical Lexicography and Lexicology (ICHLL5)*, edited by Charlotte Brewer. Oxford University Research Archive (ORA), http://ora.ox.ac.uk/objects/uuid:ef5d07d3-77fc-4f07-b13f-d4c24b4d1848.

Deutsch, Andreas. 2012. “Von der Kunst einen Schatz zu heben – Der altfriesische Rechtswortschatz und die deutsche Rechtsgeschichte.” In *Philologia Frisica anno 2008. Lêzings fan it achttjinde Frysk Filologekongres fan de Fryske Akademy op 10, 11 en 12 desimber 2008*, edited by Piter Boersma, Goffe Th. Jensma and Reinier Salverda, 26-42. Leeuwarden: Fryske Akademy.

Deutsch, Andreas. 2016. “Zur Symbiose zwischen Zettelkasten und Datenbank bei der Artikelerstellung im Deutschen Rechtswörterbuch.” In *Historische Lexikographie zwischen Tradition und Innovation*, edited by Anja Lobenstein-Reichmann and Peter O. Müller, 271-286. Berlin/Boston: De Gruyter.

Deutsch, Andreas. 2017. “Es gibt solche Schroeder - und andere. Rechtsgeschichtliche Notizen zu den Wurzeln eines Eigennamens.” In *Homo heidelbergensis: Festschrift für Klaus-Peter Schroeder zum siebzigsten Geburtstag*, edited by Pirmin Spieß, Christian Hattenhauer and Michael Hettinger, 9-24. Neustadt an der Weinstraße: Stiftung Pfälzische Geschichtsforschung.

Eichhorn, Karl Friedrich. 1834. *Deutsche Staats- und Rechtsgeschichte*, vol. I, 4th edition. Göttingen: Vandenhoeck & Ruprecht.

Ernst, Peter. 2007. “Woche und Wochentagsnamen.” In *Reallexikon der Germanischen Altertumskunde,* edited by Heinrich Beck, Dieter Geuenich and Heiko Steuer, vol. 34, 169-172. Berlin/New York: De Gruyter.

Falkson, Katharina, Ingrid Lemberg and Eva-Maria Lill. 2002. “Das Deutsche Rechtswörterbuch. Ein Wörterbuch der älteren deutschen Rechtssprache.” In *Sprache und Recht*, edited by Ulrike Hass, 355-360. Berlin/New York: De Gruyter.

Faltings, Volkert F. 2010. *Etymologisches Wörterbuch der friesischen Adjektiva*. Berlin/New York: De Gruyter.

Grimm, Jacob. 1816. “Von der Poesie im Recht.” *Zeitschrift für geschichtliche Rechtswissenschaft*, 2: 25-99.

Grimm, Jacob. 1899. *Deutsche Rechtsalterthümer*, 4th edition, supplied by Andreas Heusler. Leipzig: Dieterich.

His, Rudolf. 1901. *Das Strafrecht der Friesen im Mittelalter*. Leipzig: Dieterich.

Hofmann, Dietrich and Anne Tjerk Popkema. 2008. *Altfriesisches Handwörterbuch. Unter Mitwirkung von Gisela Hofmann.* Heidelberg: Winter.

Kirkness, Alan. 2016. “Es leben die Riesenschildkröten! Plädoyer für die wissenschaftlich-historische Lexikographie des Deutschen.“ In *Lexicographica*, 32: 17-137.

Köbler, Gerhard. 2008. “Friesisches Recht.” In *Handwörterbuch zur Deutschen Rechtsgeschichte*, 2nd edition, vol. I, 1850-1854. Berlin: Erich Schmidt.

Kroeschell, Karl A. 1995. “Germanisches Recht als Forschungsproblem.” In *Studien zum frühen und mittelalterlichen deutschen Recht*, 65-88. Berlin: Duncker & Humblot.

Künßberg, Eberhard von. 1932. “Zur Einführung.” In *Deutsches Rechtswörterbuch*, vol. I, VII-XVII. Weimar: Böhlaus Nachfolger.

Lemberg, Ingrid. 1998. “Möglichkeiten der Bedeutungserläuterung im Deutschen Rechtswörterbuch.” In *Bedeutungserfassung und Bedeutungsbeschreibung in historischen und dialektologischen Wörterbüchern*, edited by Rudolf Große, 77-85. Stuttgart/Leipzig: Hirzel.

Lemberg, Ingrid. 2007. “Lexikographie und Kulturgeschichte: 1.400 Jahre Rechtskultur im Spiegel des Deutschen Rechtswörterbuchs.” In *Sprache - Kognition - Kultur. Sprache zwischen mentaler Struktur und kultureller Prägung*, edited by Heidrun Kämper and Ludwig M. Eichinger, 151-173. Berlin/New York: De Gruyter.

Munske, Horst Haider. 1973. *Der germanische Rechtswortschatz im Bereich der Missetaten. Philologische und sprachgeografische Untersuchungen. I. Die Terminologie der älteren westgermanischen Rechtsquellen*. Berlin/New York: De Gruyter.

Nauta, Klaas. 1941. *Die altfriesischen allgemeinen Busztaxen. Texte und Untersuchungen*. Assen: Gorcum.

Nehlsen, Hermann. 1983. “Buße (weltliches Recht). II. Deutsches Recht.” In *Lexikon des Mittelalters*, vol. 2, 1144-1149. München/Zürich: Artemis.

Nehlsen, Hermann. 2008. “Reaktionsformen der Gesellschaft auf Verletzung und Gefährdung von Gemeinschaftsinteressen in Spätantike und frühem Mittelalter bei den germanischen Stämmen.” In *Das Recht und seine historischen Grundlagen, Festschrift für Elmar Wadle zum 70. Geburtstag*, edited by Tiziana J. Chiusi, Thomas Gergen and Heike Jung, 759-781. Berlin: Duncker & Humblot.

Nijdam, Han. 2008. *Lichaam, eer en recht in middeleeuws Friesland. Een studie naar de Oudfriese boeteregisters*. Hilversum: Verloren.

Nijdam, Han. 2014. “Compensating Body and Honor: The Old Frisian Compensation Tariffs.” In *Medicine and Law in the Middle Ages*, edited by Wendy J. Turner and Sara M. Butler, 25-57. Leiden/Boston: Brill.

Popkema, Anne Tjerk. 2010. “Die Bedeutung des Altfriesischen für die Erforschung des germanischen Rechtswortschatzes und vice versa.” In

Das Deutsche Rechtswörterbuch – Perspektiven, edited by Andreas Deutsch, 73-90. Heidelberg: Winter.

Popkema, Anne Tjerk. 2012. “Eine Perspektive der altfriesischen Lexikographie? Zu einem Online-Belegwörterbuch des Altfriesischen.” In *Philologia Frisica anno 2008. Lêzings fan it achttjinde Frysk Filologekongres fan de Fryske Akademy op 10, 11 en 12 desimber 2008*, edited by Piter Boersma, Goffe Th. Jensma and Reinier Salverda, 13-23. Leeuwarden: Fryske Akademy.

Richthofen, Karl Otto Johannes Theresius von. 1840a. *Friesische Rechtsquellen*. Berlin: Nicolai.

Richthofen, Karl Otto Johannes Theresius von. 1840b. *Altfriesisches Wörterbuch*, Göttingen: Dieterich.

Schumann, Eva. 2012. “Kompositionensystem.” In *Handwörterbuch zur Deutschen Rechtsgeschichte*, 2nd edition, vol. II, 2003-2011. Berlin: Erich Schmidt.

Storrer, Angelika. 2014. “Representing (Computational) Dictionaries in Hypertextual Form.” In *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Handbooks of Linguistics and Communication Science (HSK) 5/4,* edited by Rufus Hjalmar Gouws, Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand, 1244-1253. Berlin/New York: De Gruyter Mouton.

Suntrup, Rudolf. 1998. “Zahlensymbolik.” In *Handwörterbuch zur Deutschen Rechtsgeschichte*, 1st edition, vol. V, 1599-1606. Berlin: Erich Schmidt.

Vries, Oebele. 1996. “Het raadselachtige rechtswoord ‘heidenmoord’.” In *Tijdschrift voor Nederlandse Taal- en Letterkunde*, 112: 65-75.

Vries, Oebele. 2012. “Seka mit brande ende mit breke. Oudfriese terminologie met betrekking tot het rechtsinstituut ‘woesting’.” In *De taal van recht en vrijheid: studies over middeleeuws Friesland* by Oebele Vries, 284-307. Gorredijk: Bornmeer.

Vries, Oebele. 2017. “Binna skelde: zur Deutung eines Ausdrucks des friesischen Fehdewesens.” In *Amsterdamer Beiträge zur älteren Germanistik*, 77: 457-475.

Weitzel, Jürgen. 2007. “Der Strafgedanke im frühen Mittelalter.” In *Der Strafgedanke in seiner historischen Entwicklung*, edited by Eric Hilgendorf and Jürgen Weitzel, 21-35. Berlin: Duncker & Humblot.

Wetzel, Christoph. 2011. *Das große Lexikon der Symbole*, 2nd edition, Darmstadt: Wissenschaftliche Buchgesellschaft.

Wiarda, Tileman Dothias. 1805. *Asega-Buch, ein Alt-friesisches Gesetzbuch der Rüstringer, herausgegeben, übersetzt und erläutert*. Berlin/Stettin: Nicolai.

Wiarda, Tileman Dothias. 1820. *Willküren der Brockmänner, eines Freyen friesischen Volkes*. Berlin: Reimer.

Wiarda, Tilemann Dothias. 1786. *Altfriesisches Woerterbuch*. Aurich: Winter.

Wicht, Matthias von. 1746. *Das Ostfriesische Land-Recht, nebst dem Deich- und Syhlrechte, mit verschiedenen der ältesten Handschrifften zusammen gehalten und von vielen Schreibfehlern gesäubert. Durch eine nebengefügte wörtliche Uebersetzung ... erläutert*. Aurich: Tapper.

CHAPTER FOUR

JENS CHRISTIAN SVABO'S GLOSSARY: THE ORAL TRADITION AT THE BEGINNING OF FAROESE LEXICOGRAPHY

CHIARA BENATI

Abstract

The rediscovery, in the second half of the eighteenth century, of the Faroese language is strongly connected with the interest for collecting heroic ballads which had been transmitted orally over the centuries in the North Atlantic Islands. It was only thanks to the desire of collecting and making accessible this ancient cultural heritage for comparative purposes, in fact, that the Faroese language appeared in writing for the first time after it had been prohibited in schools, churches and official documents in 1536. The first collector of Faroese ballads was the Faroe-born scholar Jens Christian Svabo (1746-1824), who, not surprisingly, also compiled the first Faroese dictionary, the *Dictionarium Faeroense*, which was published about two hundred years later by Christian Matras in 1966-70, and developed the first systematic and consistent orthography of his mother tongue. In this study, I will focus on the *Collectio Vocum et Phrasium ex Carminibus Færoënsibus antiquis* (1780s), a work epitomizing per definition the close connection between Svabo's literary and lexicographic activities. In this, particular attention will be paid not only to the selection of lemmata, to the structure of their bilingual (Latin-Danish) *interpretamenta*, but also to the interaction between this glossary and the *Dictionarium Faeroense* and, through this, later Faroese lexicography.

1. Introduction to the Faroese language

Faroese is a North Germanic language spoken today by some fifty or sixty thousand persons, living in the Faroes, a group of eighteen islands situated in the North Atlantic between Scotland, Norway and Iceland, or belonging to the Faroese community in Denmark (Barnes 2005, 1794; Petersen 2005, 1839).

The colonization of the Islands dates back to the beginning of the Viking Age, when groups of Norwegian settlers landed on their coasts. Because of the origin of these immigrants, the Faroese language developed mainly from West Norwegian dialects, with which it shares many features (Thráinsson, Petersen, Jacobsen and Hansen 2004, 13). From the late Middle Ages onward, however, Faroese developed as a separate language, which survived in the Islands as spoken language even when, in 1035, the Faroes lost their independence and became tributary to Norway (Sandøy 2005, 1929) and later—after the dissolution of the Kalmar Union (1523)—was integrated in the Danish realm as an insular province.

Medieval Faroese is quite poorly documented: the main sources for this phase of the language are represented by two thirteenth-century legal documents, the *Seyðabræv* and the *Hundabræv* (Helgason 1951), containing regulations about sheep farming and dogs in the various villages on the Islands (Rindal 2002, 803). Apart from these sources, linguistic evidence from the Faroes in the Middle Ages include two runic stones (Kirkjubøur, ca. 1000 and Sandavágur, ca. 1200), whose language does not show any distinctively Faroese trait (Simonsen 1959; Brøndum-Nielsen 1923), and two series of letters: the so-called *Húsavíkarbrøvini* ("The Letters about Húsavík") dated between 1403 and 1405 (*AM fasc. 100, 1a*) and a set of letters from 1412 and 1443 (*AM fasc. 100 2a*, Jakobsen 1907, 31-53). After the Reformation (1537), Faroese completely disappeared as a written language, since every document concerning the Islands was written in Danish (Thráinsson, Petersen, Jacobsen and Hansen 2004, 369-374).

In the 1840s the introduction of compulsory schooling further increased the supremacy of the Danish language in the Islands, since every Faroese child was obliged to attend Danish classes held by Danish-speaking teachers and to study using Danish textbooks: Faroese was considered a dialect interfering in the acquisition of the Danish language and, as such, completely banned from schools, even as an auxiliary language in education (at least theoretically, since many Faroe-born teachers used their native language despite this prohibition, Vikør 2005, 2112). It took a long time for Faroese to become a subject in Faroese schools: the first step in this direction was, in 1907, the introduction of Faroese as a subject at the teacher's college in the Faroes. In 1912 a new regulation concerning primary school stated that, even though the main language of education remained Danish, Faroese could be used as an auxiliary language, if necessary. Eventually Faroese was made a compulsory subject (first only oral, from 1920 also written), but not until 1938 did it become the general language of instruction in Faroese schools (Vikør 2005, 2112).

This increased importance of the Danish language in the Faroese society led to a reaction aimed at promoting the use of the Faroese language in all communicative contexts. This emancipation process culminated in 1948, when Faroese was recognized as the main language in the Islands with the Home Rule Act (*Lov om Færøernes Hjemmestyre*), even though Danish continued to have legal status and to be widely used in literature, media and in various other pragmatic contexts.[1]

2. Jens Christian Svabo and the rediscovery of the Faroese language

As mentioned before, the Reformation represents a turning point in the linguistic history of the Faroes: unlike Iceland, the Faroes did not have their own translation of the Bible. The Danish Bible was introduced instead, and Danish became the language of the Faroese Church. Being used by both the political and the religious authorities, Danish was then the only written language in the Islands. As observed by Hagström, "for hundreds of years the linguistic situation was a kind of diglossia" (Hagström 1985, 189), where Faroese only survived as a bunch of spoken dialects without any standard norm, while Danish was the official language, used in writing and in every official circumstance. Among themselves the Faroese continued, however, to speak Faroese, since Danish remained a mere foreign means of communication: it was the language of the clergy and the officials, socially quite distant from the local rural population (Hagström 1985, 189).

The rediscovery of Faroese and, consequently, its development as written language is strongly connected with the interest for collecting the heroic ballads which had been transmitted orally over the centuries in the North Atlantic Islands. It was only thanks to the desire of collecting and making accessible this ancient cultural heritage for comparative purposes, in fact, that the Faroese language appeared in writing for the first time after the Reformation. The first collector of Faroese ballads was the son of the Danish pastor of the island of Vágar, Jens Christian Svabo (1746-1824)

[1] About language the 1948 Home Rule Act says: "Faroese is recognized as the main language, but Danish is to be learned well and thoroughly, and Danish may be employed in official contexts as well as Faroese. When law suits are forwarded to a higher court of appeal [which is situated in Denmark] all documents in Faroese should be followed by a Danish translation." (Vikør 2005, 2112). On this, see also Benati 2009, 192-193.

who, in the 1770s, began to write them down and also compiled the first Faroese dictionary (Thráinsson, Petersen, Jacobsen and Hansen 2004, 374; Dalen 2005, 1415; Considine 2017, 217; Haugen 1968, 159).

Since Faroese had not been written down for centuries, Svabo had to develop a spelling system for it. This first orthographic proposal for the Faroese language, which was based on Svabo's own dialect (that of the island of Vágar), can be described as 'phonetic', since it reflects many of the typical features of Faroese phonology (e.g. diphthongization, dissimilation, palatalization and *skersping*, Schulte 2005, 1089-1092) and, therefore, basically shows a one-to-one relation between graphemes and phonemes (Thráinsson, Petersen, Jacobsen and Hansen 2004, 374-376; Haugen 1968, 161). Despite the evident advantages in terms of ease of learning for both native children and foreigners, Svabo's orthography did not establish itself as standard, since the spelling of Modern Faroese is an etymological one. This was introduced, in the nineteenth century, by Venceslaus Ulricus Hammershaimb (1819-1909), who, through this, wanted to give his language historical dignity, by stressing its relationship with the Old Norse and Icelandic traditions [2] (Thráinsson, Petersen, Jacobsen and Hansen 2004, 380-392; Haugen 1968, 162-63; Vikør 2002, 8; Dalen 2005, 1415).

Svabo's lexicographic activity can be considered his "major life preoccupation" (Haugen 1968, 160): already in his early twenties, as a student of economics and natural history at the University of Copenhagen, he started collecting and translating Faroese words. The first recension of his Faroese-Danish-Latin dictionary, entitled *Forsøg til en Ordbog eller Ordsamling i det færøeske Sprog*, is dated June 18, 1773. This is transmitted in four slightly diverging manuscripts and includes some 4,900 entries (Haugen 1968, 160; Considine 2017, 217-218). In the early 1780s Svabo was given government sponsorship to travel back to the Islands and run a survey about Faroese economy. While conducting the fieldwork for this research, he became more and more interested not only in the language of the Islands, but also in their oral tradition and, in particular, in the medieval ballads which, as a single case in all Scandinavia, were still sung and danced and played a fundamental role in marking festivities and special occasions

[2] The difference between the two orthographic systems is evident when comparing the spelling of words like "pleasure" [gle:j̥ɪ] *Gleji* (Svabo's spelling, Sv.) vs. *gleði* (Hammershaimb's spelling, Hm.), "high" [hø:an] *Høvan* (Sv.) vs. *høgan* (Hm.), "blue whale" [ɹai:jʊɹ] *Rojur* (Sv.) vs. *royður* (Hm.), "few" [fɔa:vʊɹ] *Faavur* (Sv.) vs. *fáur* (Hm.), "cow" [k^{h}ɪgv] *Kygv* (Sv.) vs. *kúgv* (Hm.), "white" [kvʊi:tʊɹ] *Kvujtur* (Sv.) vs. *hvítur* (Hm.), "red" [ɹɛi:jʊɹ] *Rejur* (Sv.) vs. *reyður*, or "to call" [k^{h}adla] *Kadla* (Sv.) vs. *kalla* (Hm.).

for the Faroese society. On his research tour to the Faroes, in fact, not only did he expand the first version of his dictionary by adding about half as many more entries, but he also prepared a large ballad manuscript collection containing fifty-four ballads. This was supplied with a glossary of difficult and unusual words appearing in the texts, which will be dealt with in this study.

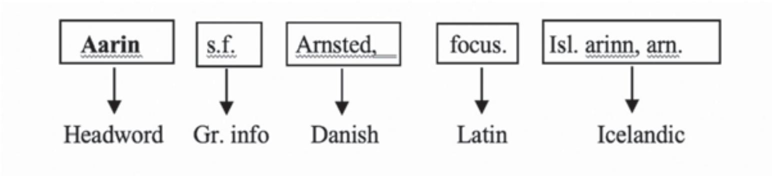
In his last years in the Faroes, Svabo, who had no access to his previous manuscripts, continued his lexicographic work, starting from scratch and compiling a new dictionary containing some 7,500 entries (Haugen 1968, 160).[3] The main motivation behind Svabo's collecting effort was the gloomy vision of the extinction of Faroese language and literature, which he, therefore, wanted to record, in order to let people know what the language and literature looked like, once they had definitively died out (Considine 2017, 218).

3. The *Collectio Vocum et Phrasium ex Carminibus Færoënsibus antiquis* (Svabo's Glossary)

Compiled in the 1780s along with the collection of ballads, to which it should constitute an addendum, Svabo's glossary is transmitted in two manuscripts written by Nicolai Mohr (1742-1790): Copenhagen, Kongelige bibliotek, Ny kgl. Saml. 1953 4to (G_1) and Ny kgl. Saml. 1951 4to (G_2) (Svabo 1943, 3), which have both been used in Christian Matras' edition (1943). The glossary, which is based on the texts of the ballads included in Svabo's own collection, represents the first attempt to translate and explain the antiquated and uncommon words and phrases used in Faroese oral tradition, which are rendered in Danish and Latin and—usually—compared with their Icelandic equivalent.

On the whole, the glossary includes some eight hundred entries ordered alphabetically and divided into twenty-one letter-sections. The minimal structure of the single entries can be described as follows: headword, word class and other grammatical pieces of information (gender for nouns, inflection details for verbs), Danish, Latin and—if present—Icelandic equivalent(s). See, for example, the entry for *aarin* (*árin*) "fireplace" (Svabo 1943, 9):

[3] He did not have access to his previous manuscripts because they were in the hand of Nicolai Mohr, with whom he had collaborated; Mohr died in 1790 but the manuscripts did not return to Svabo.



In some entries, this basic structure is expanded with further elements, such as compounds and collocations in which the headword appears, or the indication of the source—i.e. the ballad—in which it is used. See, for example, the case of *aasir* (*ás*, *ásir*) "columns" (Svabo 1943), where the actual lemma is in the plural form, whereas the singular form is provided (and translated differently) only later in the entry, after the compound *hura-aasir* "doorposts" and its Danish and Latin translations:

> **Aasir** s. pl. f. Stolper, columnæ. Hura-aasir, Dør-Stolper, columnæ forium. Hodie est Aas, Stang, pertica. Isl. ásar m. pl., trabes s. tigni horizontales.

In this case, Svabo also underlines the differences between Faroese and Icelandic in terms of both grammatical gender and meaning: while, in fact, in Faroese the term is feminine and can be rendered in Latin as *columnae* "columns", in Icelandic it is masculine and translated as *trabes* "beams".

A similar comparative metalinguistic consideration can be found in the entry for the verb *braaä* (*bregða*) "to hurry, to spin" (Svabo 1943, 9):

> **Braaä** v. n. skynde sig, svinge sig til, vibrare, festinare. Isl. bregda. Braa in carmine Grimars Kveaï et alibi invenitur, ubi verisimiliter imperfectum est verbi braaä. Hann braa see uj Ørna lujkji, Han paatog sig hastigen en Ørns Skikkelse, celeriter se in aquilam l. aquilæ speciem mutavit (Grim. Kv.) Isl. Hann brá ser í arnarlíki. Braa preteritum verbi Islandorum, bregd, bregda, cito se vertere, nam præter hoc tempus vix in idiomate Færöensi exstat. Han braa sujnum Svøri, Han svingede sit Sværd, vibravit gladium. Isl. brá sínu sverdi.

In this case, the very attribution of meaning to a series of passages in Faroese ballads containing the form *braa* (*brá*) is determined by the comparison with Icelandic, where *brá* is the preterite of the verb *bregða* "to draw", which can, therefore, be reconstructed for Faroese as well.

Given the transmission context of Faroese ballads and their centuries-old exclusively oral tradition, it is no surprise that they preserved terms and words which were no longer used elsewhere or understood at the time Svabo first transcribed them. This particular situation is often reflected in the *interpretamenta* of his glossary, where he tries to reconstruct the meaning of a term on the basis of the comparison with Icelandic, comments on the possible corruption of certain passages, trying to reconstruct their original diction, or accounts for the difference in meaning between contemporary Faroese and the ballads. See, for example:

> **Higgjubekkur**, alii canunt Kalbekkur, sed utrumque vocabulum vix invenitur, nisi unico loco carminis Aalvur Kongur, ubi transtrum significare videtur, aut forte potius Udkikssted, sedem speculatoris in navi, ab Isl. hyggia, speculari; et Kalbekkur, sedem in navi vento maxime oppositam (Svabo 1943, 37).
>
> **Leusnari** in carmine Ljeumunar invenitur: Ee lovi tea uj Leusnari nej, sed hoc valde corruptum, et ita corrigendum: Ee lovi tee Leusnarin aj. Isl. Eg lofa þig Lausnarinn æ, Jeg priser Dig ævigen min Frelser, te, Salvator! æternum laudo (Svabo 1943, 47-48).
>
> **Rust** s. f. hodie potissimum significat dorsum montis acclivis, sed in carmine Gongu Rølvur, ubi exstat: Vervur o Vatn eav Rustun koom, significat tantum, et højt Field, montem altiorem. Fortè Isl. reist, erectio, aut antiqu. raustn, altitudo, sublimitas, hujus vocis origo sit, a reisa, erigere, surgere (Svabo 1943, 62).

As these examples clearly show, Svabo's glossary contains the earliest historical semantic considerations about Faroese vocabulary and its diachronic evolution from the past—represented here by the diction transmitted orally in the ballads—and the late eighteenth-century spoken language which Svabo called 'mother-tongue'. In this respect, the insertion of Icelandic equivalents to the headwords seems to have a specific function, which cannot be assimilated to the one of the other two languages used in the glossary, i.e. Danish and Latin.

In the late eighteenth-century Latin was not only the language of education and culture, but also the lexicographic metalanguage *par excellence*. For this reason, Svabo chose it to provide grammatical details, explanations and a universally understandable translation of the headwords included in the glossary as well as in his dictionary.

Danish was Svabo's ideal language as opposed to the corrupted form of contemporary Faroese, the one he hoped Faroese people would finally learn to speak in its purest form.[4] Apart from these ideal considerations,

[4] See Svabo 1966-1970, 2: xv: "Jeg kan desuden ikke andet end forestille mig det som en langt større Fuldkommenhed, at Kolonien, saalænge den med Dannemark tilbeder en Gud, knæler for en Konge, ogsaa talte med samme Tunge som det." See also Haugen 1968, 161 and Barnes 2005, 2008.

however, the choice of translating Faroese words and phrases into Danish also had a very practical justification, since the glossary's intended audience—educated Faroese and Danish people interested in the Islands and in their culture—read and wrote in Danish.

Deciding to insert Icelandic equivalents to the headword, on the other hand, Svabo did not aim at reaching a wider audience (Icelandic was—and still is—not exactly a well-known language!) or target specifically Icelandic-speaking readers, but he rather wanted to put in plain sight the main epistemological instrument that he had used to understand and explain the meaning of the antiquated and—sometimes—corrupt words of the ballads which he had recorded during his tour in the Islands. When his own linguistic proficiency or that of his informants were not sufficient, he tried to reconstruct their meaning on the basis of the comparison with Icelandic, being the most direct descendant of the glorious Old Norse.

In this, Svabo unconsciously laid the basis for the later development of the Faroese language, which is ideally oriented to Iceland and Icelandic, as the adoption of Hammershaimb's etymological spelling clearly demonstrates. Strongly inspired by the mid-nineteenth-century National Romantic and by the contemporary cultural debate, Hammershaimb legitimated the claim for dignity of Faroese on its being an independent language, which had preserved its Old Norse traits (Thráinsson, Petersen, Jacobsen and Hansen 2004, 382). Moreover, thanks to its closeness to Old Norse and Icelandic, Hammershaimb's etymological orthography would give non-Faroese readers easier access to the rich oral tradition, which was being written down in those decades (Barnes 2005, 2008) and which began to draw attention to the Islands and their language.

In this respect, Svabo's collection of ballads is at least as important for the future development of the Faroese language as his lexicographic activity and the two are so strongly intertwined that it is extremely difficult to distinguish between them. For this reason, it is particularly interesting to investigate the interaction between the glossary and Svabo's main lexicographic effort, the *Dictionarium Faeroense*, in order to ascertain if the antiquated and unusual words which he had isolated in the corpus of ballads collected during his tour of the Islands were inserted in his dictionary as part of the ca. 2,500 lemmata which were added to its first version. An analytic comparing of the presence of the entries included in the glossary with both Svabo's *Dictionarium Faeroense* and the *Føroysk Orðabók* (the first Faroese monolingual dictionary of 1998[5]) will outline the role of this small

[5] See Poulsen, Simonsen, Jacobsen, Johansen and Hansen 1998; see also Barnes 2005, 2010.

lexicographic work in the later standardization of Faroese and in the development of a Faroese lexicographic tradition.

Within the time and space constraints of a study like this, this analysis will be conducted on a sample of lemmata constituted by the *A*- and *N*-sections of the glossary, the representation of which in the two major lexicographic works will be systematically verified.

Svabo's Glossary	*Dictionarium Faeroense*[6]	*Føroysk orðabók*
Aalbrinjavur	Aalbrinjavur S^{1} *M*	albrynjaður
Aarin	Aarin S^{1} *M*	árin^{2} (*kvæð.*)
Aasir	Aas S^{1-6} *M*	ás, ásur
Aast	Aastir S^{1}	ást
Aastar-Bragd *Vi hajtun Aastar-bragdi*	-	Ástarbragd (*kvæð., skaldsk.*) *Við heitum ástarbragdi*
Adroj	-	-
Agnar-Oor	Agnar-Oor *M*	agnarorð
Aja	Aja S^{1-6} *M*	eiga^{2}
Ajin	-	eiga^{1}
Ajn	Ajn S^{1-6} *M*	ein
Ajns Mans Gjeer *Tea eer ikkji ajns Mans Gjeer*	-	einsmansgerð *tað er ikki ~*
Ajti	Ajti S^{1} *M*	heiti
Ajtur	Ajtur S^{1-4} *M*	eitur
Ajtur-bland *Lejggja mengari Kjempu firi Ajtur-bland*	-	eiturbland *leggja mangari kempu fyri ~*
Aldans-Fiskur	Aldans-fiskur *M*	aldansfiskur (*kvæð.*)
Aldur	Aldur s.m. Alder, ætas, Isl. id. $S^{2\ 3\ 5\ 6}$ *M*	aldur
Alvi	Alvi S^{1}	alv
Angur	-	angur (*kvæð.*)

[6] The *Dictionarium Faeroense* is quoted here on the basis of its 1966 edition by Christian Matras (Svabo 1966-1970). The indication of the manuscript versions of the dictionary should be read as follows: S^{1} = Copenhagen, Kongelige bibliotek, AM 971, 4to, which constitutes the main witness of the 1800 redaction; S^{2-6} = the manuscripts transmitting the first redaction from the 1770s, *M* = Copenhagen, Kongelige bibliotek, Ny kgl. Saml. 1287 fol., the main witness of the 1780s.

Ann-nes	Annees *M*	and-nes
Ansa	Ansa S^{1-6} *M*	ansa
At	At *M*	ætt
Atlajä, l. Atleaa	-	ætlan
Avdeavur	Av-deavur *M*	avdagur
Axla	-	aksla (*kvæð.*)

Naaïr	Naaï S^{1} *M*	náð
Naat-beul	-	náttból
Naatt *Aa miäri Naatt*	Naat S^{1-6} *M*	nátt *á miðjari ~*
Najpa	-	neipa (*kvæð.*)
Najsa *Vinna ajnun Najsir*	Najsa S^{1} *M*	neisa *vinna e-m neisur*
Nejgiifta	Nejgjifta *M*	-
Nemmast	-	-
Nevndur firi	Nevndur firi *M*	-
Njeusnari	-	njósnari
Norna	-	norn
Nossu-gaava	-	-
Nuja	-	níga
Nuja *Aa ajn*	-	níða (*kvæð.*) *~ á e-n*
Nujings-majn *Vin naajnun Nujingsmajn*	-	níðingsmein *hann vann honum ~*
Nujta *in eat Hjarta*	Nujta S^{1} *tea nujtur uj Hjarta*	níta *tað neit henni í hjartað*
Nujtur *Draangur*	Nujtur *M* *Draangur*	nýtur (*kvæð.*) *tann hin nýti drongur*
Nunna	Nunna $S^{2\ 3\ 4\ 6}$ *M*	nunna
Nygv	Nygv *M*	nýgv

As the chart above clearly shows, the sample of lemmata from Svabo's glossary is differently represented in the other two lexicographic sources taken into consideration for this study. The most striking datum emerging from this comparison is that these words are more likely to be found in the 1998 *Føroysk Orðabók*, than in Svabo's own dictionary.

As for the *Dictionarium Faeroense*, only part of the entries from the glossary have been included in it. Thanks to the philological approach of Matras' edition of the dictionary, it is possible to follow the exact chronology of its versions and integrations and to verify the assumption that they were, for the largest part, inserted in the manuscripts *S¹* and *M*, that means either contextually to their first recording, i.e. during Svabo's tour of the Islands, or in the third and last version of the dictionary, the one he compiled from scratch after 1800, after he had moved back to the Faroes. Only in a few cases, on the other hand, did Svabo insert in the glossary the lemmata which were already present in his 1770s manuscripts of the dictionary (see, for example, *Aas, Ansa, Naat, Nunna* etc.). In these cases, he usually provided a new meaning for the term which did not correspond to the one(s) given in the manuscripts of the dictionary. This is, for example, the case of the entry for the masculine *Aldur*, which appears in this way in the glossary:

> **Aldur** s. m. Abildgaard, Abild-Have, pomerium. Ab Isl. apalldr, pomus, (Svabo 1943, 10)

whereas in the dictionary we find this:

> **Aldur** s. m. Alder, ætas. Isl. id. (Matras 1966, 19)

We can imagine that Svabo first discovered the use of the term *aldur* in the meaning of "(crab) apple orchard" while travelling around the Islands and recording the ballads and that he decided not to include it in the dictionary because he considered this meaning too rare or too specific of poetic texts.

As we have seen, in fact, Svabo was quite selective when it came to including in his larger lexicographic work the words he had isolated in his collection of ballads. Only 24 (60%) of the selected lemmata from the *Collectio vocum et phrasium* could be found in the *Dictionarium* Faeroense. Moreover, a comparison of the entries in the glossary and in the dictionary seems to hint at the systematic omission of two categories of words: 1. compounds, which Svabo must have considered occasional, of uncertain interpretation or whose meaning could be inferred on the basis of their components (see *Aastar-Bragd, Ajn Mans Gjer, Ajtur-bland, Naat-beul, Nossu-gaava, Nujings-majn*) and 2. simplices, which, according to Svabo's notation, only occur in the text of one ballad (see *Adroj* in *Hearald-skjøldur*; *Ajin* in *Grimur aa Miälnesi*; *Najpa* in *Aalvur kongur*). For other entries, such as *Axla, Atleaa*, etc., the reason for their non-inclusion in the dictionary does not appear so evident, but we can assume their exclusion is again due to either the uncertainty of their interpretation or to their status of *hapax*, possibly originating from a corrupt transmission.

In the *Føroysk Orðabók*, on the other hand, we find a correspondence for a large majority of the selected lemmata (35 out of 40, 87.5%). This datum is even more striking if we consider that we can exclude a direct dependence from Svabo's glossary, which does not belong to the sources cited by the authors of the Faroese monolingual dictionary, where the only works of Svabo's mentioned are the *Dictionarium Faeroense* and the *Indberetninger fra en Reise i Færøe 1781 og 1782* (Poulsen, et. al, 1998, 24-26). Yet, the influence of the glossary is undeniable: many of the monolingual *interpretamenta* include exactly the same collocations quoted by Svabo, see, for example, *Vi hajtun Aastar-bragdi – við heitum ástarbragdi*; *Lejggja mengari Kjempu firi Ajtur-bland – leggja mangari kempu fyri ~*; *Aa miäri Naatt - á miðjari ~*. This very fact suggests that—possibly also thanks to Svabo's ballad collection—these words and phrases had reinforced their status within the arising corpus of written Faroese, so that they became consolidated enough to be taken into consideration for inclusion in the first monolingual dictionary of the Faroese language. The same reasoning can be applied to all those lemmata appearing in the *Føroysk orðabók* without any indication of their being used exclusively in ballads (*kvæð.*) or in poetry in general (*skaldsk.*), since the absence of such information leads to the assumption that the term can also be used in other contexts.

4. Conclusion

Jens Christian Svabo is universally recognized as one of the 'fathers' of the Faroese language and is usually remembered as either the first collector of the *kvæðir* (and creator of the first systematic Faroese orthography) or as the author of the first Faroese dictionary. Nevertheless, the close connection between these two fields of interest of his has never been investigated in depth. If, on the one hand, in general terms the relation between the very survival of Faroese, as opposed to Orkney and Shetland Norn, and the tradition of the *kvøldseta* (literally "evening sitting"),[7] which allowed the preservation of a common historical, literary and linguistic inheritance, has

[7] *Kvøldsetur* were the evening assemblies the Faroese families used to have in the long, dark and cold winter nights. During these gatherings, women spun or knitted and men played cards, while stories were told and ballads sung in turn. In this way, this common inheritance remained an important part of Faroese everyday life up to the end of the nineteenth century when economic change rendered *kvøldsetur* obsolete as productive occasions.

often been highlighted, on the other, the importance of oral tradition for metalinguistic analysis and lexicography has never been scholarly described. Specifically, neither the interaction between Svabo's own literary and lexicographic activities nor that between his major lexicographic achievement—the *Dictionarium Faeroense*—and the glossary in addendum to his collection of ballads had ever been really analyzed.

In this study I have focused on the glossary, a work ideally combining the activity of Svabo the collector and Svabo the lexicographer. The analysis of this almost completely neglected lexicographic work has shown that, far from being merely functional to the understanding of the ballads, it constitutes a valuable source for the semantic diachronic development of Faroese vocabulary. In a transmission context like that of Faroese ballads, which had survived for centuries exclusively in oral form, the meaning of some words had often to be reconstructed on the basis of a comparison with Icelandic, the most direct and incorrupt descendant of Old Norse. In this respect, the insertion of Icelandic equivalents to the headwords included in the glossary should not be seen as a third translation along with the Latin and the Danish ones, but rather as the exposition of the main epistemological instrument employed by Svabo for an in-depth understanding of his mother tongue.

Moreover, the cross check of the presence of a sample of entries from the glossary in the various versions of the *Dictionarium Faeroense* has allowed outlining Svabo's idea of interaction between the two lexicographic works: antiquate poetic lemmata from the ballads were not included automatically and indiscriminately in the second and third version of the dictionary, but only if they were not considered occasional coinings or *hapax*, possibly resulting from a corrupt transmission.

Despite their exclusion from the *Dictionarium Faeroense*, however, many lemmata from Svabo's glossary have survived in later Faroese lexicography, as witnessed by their presence in the *Føroysk Orðabók*, the 1998 monolingual Faroese dictionary. The insertion of these words and phrases in the *Føroysk Orðabók* could possibly be influenced by the wish of compiling a large reference dictionary, which could make a statement about the dignity of the Faroese languages. Nevertheless, it suggests that, in the decades after the compilation of Svabo's collection of ballads, these words must have started being used in other contexts than the poetic one and have become consolidated enough to be included in a dictionary aimed at becoming the standard reference work for Faroese. This provides further evidence for the thesis that Svabo, linguist and lexicographer, cannot be distinguished from Svabo, collector of ballads, and that oral poetry played a fundamental role not only in the survival of the Faroese language, but also in the long emancipation process, which brought the low variety of a centuries-old diglossia to become the main and official language of the North Atlantic Islands.

Bibliography

Barnes, Michael P. 2002-2005. "Social stratification in the present-day Nordic languages V: Faroese." In *The Nordic Languages. An International Handbook of the History of the North Germanic Languages*. 2 vols., edited by Oskar Bandle, Kurt Braunmüller, Ernst Håkon Jahr, Allan Karker, Hans-Peter Naumann and Ulf Telemann, 2: 1794-1796. Berlin and New York: Walter de Gruyter.

Benati, Chiara. 2009. "Faroese: A National Language under Siege?" In *Rights, Promotion and Integration Issues for Minority Languages in Europe*, edited by Susanna Pertot, Tom M.S. Priestly and Colin H. Williams, 189-196. Houndmills: Palgrave Macmillan.

Brøndum-Nielsen, Johannes. 1923. "Sandevåg-runestenen (Vågo-Færøerne)." *Aarbøger for nordisk Oldkyndighed og Historie*, 3(13): 111-122.

Considine, John. 2017. *Small Dictionaries and Curiosity. Lexicography and Fieldwork in Post-Medieval Europe*. Oxford: Oxford University Press.

Dalen, Arnold. 2002-2005. "Sources of written and oral languages in the 19th century." In *The Nordic Languages. An International Handbook of the History of the North Germanic Languages*. 2 vols., edited by Oskar Bandle, Kurt Braunmüller, Ernst Håkon Jahr, Allan Karker, Hans-Peter Naumann and Ulf Telemann, 2: 1406-1419. Berlin and New York: Walter de Gruyter.

Hagström, Björn. 1985. "The development of Scandinavian languages as national languages." In *Entstehung von Sprachen und Völkern, glotto- und ethnogenetische Aspekte europäischer Sprachen. Akten des 6. Symposions über Sprachkontakt in Europa, Mannheim 1984*, edited by Per Sture Ureland, 181-192. Tübingen: Niemeyer.

Haugen, Einar. 1968. Review of Jens Christian Svabo, *Dictionarium Færoense*, edited by Christian Matras, vol. I. *Scandinavian Studies,* 40(2): 159-163.

Helgason, Jón. 1951. "Kongsbókin úr Føroyum." *Utiseti,* 6: 101-122.

Jakobsen, Jakob. 1907. *Diplomatarium Færoense. Føroyskt Fodnbrævasavn. Miðaldarbrøv upp til trúbótarskeiðið við søguligum ransóknum*. Tórshavn: H.N. Jacobsen.

Petersen, Hjalmar P. 2002-2005. "Trends in the linguistic development since 1945 V: Faroese." In *The Nordic Languages. An International*

Handbook of the History of the North Germanic Languages. 2 vols., edited by Oskar Bandle, Kurt Braunmüller, Ernst Håkon Jahr, Allan Karker, Hans-Peter Naumann and Ulf Telemann, 2: 1839-1843. Berlin and New York: Walter de Gruyter.

Poulsen, Jóhan Hendrik W., Simsonsen, Marjun, Jacobsen, Jógvan í Lon, Johansen, Anfinnur, and Svabo Hansen, Zakaris. 1998. *Føroysk Orðabók*. Tórshavn: Føroya Fróðskaparfelag og Fróðskaparsetur Føroya.

Rindal, Magnus. 2002-2005. "The history of Old Nordic manuscripts II: Old Norwegian (incl. Faroese)." In *The Nordic Languages. An International Handbook of the History of the North Germanic Languages*. 2 vols., edited by Oskar Bandle, Kurt Braunmüller, Ernst Håkon Jahr, Allan Karker, Hans-Peter Naumann and Ulf Telemann, 1: 801-808. Berlin and New York: Walter de Gruyter.

Sandøy, Helge. 2002-2005. "Sociolinguistic structures chronologically IV: Icelandic and Faroese." In *The Nordic Languages. An International Handbook of the History of the North Germanic Languages*. 2 vols., edited by Oskar Bandle, Kurt Braunmüller, Ernst Håkon Jahr, Allan Karker, Hans-Peter Naumann and Ulf Telemann, 2: 1923-1932. Berlin and New York: Walter de Gruyter.

Schulte, Michael. 2002-2005. "Phonological developments from Old Nordic to Early Modern Nordic I: West Scandinavian." In *The Nordic Languages. An International Handbook of the History of the North Germanic Languages*. 2 vols., edited by Oskar Bandle, Kurt Braunmüller, Ernst Håkon Jahr, Allan Karker, Hans-Peter Naumann and Ulf Telemann, 2: 1081-1096. Berlin and New York: Walter de Gruyter.

Simonsen, Marie Ingerslev. 1959. "The Kirkjubø Runic Stone." *Acta philologica scandinavica*, 24: 107-124.

Svabo, Jens Christian. 1943. *Svabos Glossar til færøske visehaandskrifter*, edited by Christian Matras. Copenhagen: Bianco Lunos Bogtrykkeri.

Svabo, Jens Christian. 1966-1970. *Dictionarium faeroense*, edited by Christian Matras. 2 vols. Copenhagen: Munksgaard.

Thráinsson, Höskuldur, Petersen, Hjalmar P., Jacobsen, Jógvan í Lon, and Svabo Hansen, Zakaris. 2004. *Faroese. An Overview and Reference Grammar*. Tórshavn: Føroya Fróðskaparfelag.

Vikør, Lars S. 2002-2005, "The Nordic language area and the languages in the north of Europe." In *The Nordic Languages. An International Handbook of the History of the North Germanic Languages*. 2 vols., edited by Oskar Bandle, Kurt Braunmüller, Ernst Håkon Jahr, Allan

Karker, Hans-Peter Naumann and Ulf Telemann, 1: 1-12. Berlin and New York: Walter de Gruyter.

Vikør, Lars S. 2002-2005. "Bilingualism in Schleswig, Finland, North Sweden, Northern Norway, the Faroe Islands, and Greenland." In *The Nordic Languages. An International Handbook of the History of the North Germanic Languages*. 2 vols., edited by Oskar Bandle, Kurt Braunmüller, Ernst Håkon Jahr, Allan Karker, Hans-Peter Naumann and Ulf Telemann, 2: 2105-2114. Berlin and New York: Walter de Gruyter.

CHAPTER FIVE

A MINOR PERIOD OF A MAJOR LANGUAGE: LEXICOGRAPHICAL APPROACHES TO LATE OLD ENGLISH

RACHEL FLETCHER

Abstract

This article examines from a historical perspective the creation and handling of late Old English as a marginal category within a wider lexicographical tradition. Over the past five centuries, lexicographers of Old English have shown an increasing interest in describing certain vocabulary, texts or manuscripts as 'late', setting them apart from classical, pre-Conquest Old English as a marked category.

I examine seven major dictionaries of Old English from the seventeenth century to the present day: those of William Somner (1659), Edward Lye and Owen Manning (1772), Joseph Bosworth (1838), Joseph Bosworth and T. Northcote Toller (1882-98), J.R. Clark Hall (1894), and Henry Sweet (1896), and the University of Toronto's *Dictionary of Old English* (Cameron et al. 2018). These dictionaries are used to illustrate the varying descriptive approaches that establish the concept of late Old English as a distinct yet subordinate category. Although the majority of these works identify a category of late Old English (or something functionally similar) as an object of linguistic interest, they differ widely in what material is considered late, how this identification is expressed and how systematically it is applied. Nevertheless, it is possible to identify shared themes that reflect a continuity in Old English scholars' attitudes towards lateness.

1. Introduction: Late Old English as a minor period

'Lateness of composition clearly must result in making a text inadmissible for a dictionary of Old English' (Stanley 1985, 331). This statement, taken from a volume of collected essays on the topic of problems in Old English lexicography, may well appear to be a matter of common sense. From the point of view of lexicographical history, however, the consensus that lateness is an issue meriting particular comment only emerged gradually.

The characterization of the issue as a matter of 'lateness' reveals certain assumptions about the division of language history. Late Old English is presented as subordinate to an unmarked standard form of Old English, rather than being analyzed on its own terms or in relation to the Middle English by which it is followed. As will be seen, this subordination is evident in many dictionaries of Old English, with late material being included as a marked category.

Within the wider field of general Old English studies, there has been relatively little agreement on how 'late' Old English should be identified, or indeed on how best to identify the end of the Old English period, with articles such as those by Malone (1930), Kitson (1997) and Lutz (2002) demonstrating a variety of approaches. Likewise, different lexicographers have identified the sub-period of late Old English in various ways, although the English used in (approximately) the century and a half following the Norman Conquest of 1066 has generally been treated as the prototypical core.

Furthermore, it has been observed that texts considered to be late have often received less attention in scholarship, distorting understanding of this period and its language and literature (e.g. Treharne 2012). When such texts, and the language they are written in, are discussed, the sense of late Old English as being a subordinate period may also tie into ideas of linguistic purity; late Old English is portrayed as the corruption of a standard. Such concerns are evident in some of the earliest scholarship on Old English.[1] Attitudes such as these could in turn be taken as justifications for the continued neglect of late Old English in scholarship.

Taken together, the factors of subordination, lack of consensus and paucity of scholarship led me in the title of this article to characterize late Old English as a 'minor' period, despite the dominance of Old English as a whole. Although these are observations about general trends in Old English

[1] See for instance Cain (2010) on anxieties about language contact. At least one Early Modern editor, William L'Isle, even went so far as to emend thirteenth-century material to bring it closer to standard Old English; see Pulsiano (2000, 187-191) and Lee (2000, 225-231).

scholarship, they can also be seen specifically in the discipline of historical lexicography. It is very rare to find a dictionary dealing with late Old English in its own right; the closest is perhaps Herbert Coleridge's *A Glossarial Index to the Printed English Literature of the Thirteenth Century* (1859). However, as the title suggests, it is not a full dictionary, and was originally conceived as a tool to be used in a larger lexicographical project, the *Oxford English Dictionary.* In any case, the thirteenth-century material it covers would be considered by many as too late to count even as late Old English, and this appears to be the approach taken by Coleridge, who sees the mid-thirteenth century as marking the beginning of 'English literature, as distinguished from Semi-Saxon' (Coleridge 1859, iii).

More usually, late Old English is treated as a marked category within dictionaries of Old English. Of course, since the boundaries between linguistic periods are neither objective nor absolute, the same span of linguistic history that is treated in dictionaries of Old English as late may also be treated by lexicographers of Middle English as early Middle English. In this article, however, I restrict my attention to how lateness is imagined and represented by lexicographers of Old English.

In this article, I do not intend to propose another, new set of criteria for defining a late Old English period; neither do I intend to critique existing dictionaries of Old English on their coverage of material that might be considered late. Instead, I mean to trace how the concept of late Old English as an identifiable—yet distinctively 'minor'—category has developed through the history of Old English lexicography. What traditions of scholarship lie behind the current lexicographical convention of late Old English as a minor period, subordinate to the unmarked standard form of the language?

To answer these questions, I examine seven major dictionaries of Old English produced from the seventeenth century to the present day.[2] Although they differ in their understanding of the chronological scope of the Old English period, their terminology and the methodologies and aims of their lexicographers, it is still possible to trace developments in the representation of late Old English in dictionaries, though these are not necessarily perfectly linear.

[2] Bailey (1990, 1437-1451) provides an overview of the lexicography of Old English that places the dictionaries discussed in this article in a wider historical context, although its date of publication means that it does not cover some recent developments, particularly with respect to the Toronto *Dictionary of Old English.*

2. Early approaches: the *Dictionarium Saxonico-Latino-Anglicum* (1659) and the lack of a transition period

The first dictionary of Old English to be published was William Somner's *Dictionarium Saxonico-Latino-Anglicum* (1659).[3, 4] Unlike subsequent dictionaries of Old English that are comparable in scope and detail, Somner's work does not appear to recognize late Old English as a noteworthy or distinctive category. To a certain extent, it could be argued that this is the result of Somner's general lack of attention to diachronic variation within Old English. He, and his contemporaries, had access to fewer manuscripts than did later scholars, and had fewer resources at their disposal to assess the relative ages of these manuscripts; it was thus natural for him to view Old English as a homogeneous category and—in spite of his awareness that it was spoken over several hundred years—to describe it from a synchronic perspective.

However, Somner does at times show an awareness of time depth and language change in the period covered by his dictionary. When, in his preface, he discusses the challenges that orthographical variation presents to the lexicographer, he comments in passing that such variation may occur 'pro varia scilicet vel ævi vel loci dialecto' (Somner 1659, *Ad Lectorem* §14).[5] What distinguishes Somner's attitude towards diachronic variation, as expressed in his preface, from the attitudes of later lexicographers is the fact that, when he identifies a specific text as being linguistically unusual on account of its age, this anomalousness is associated not with lateness but with earliness. Elsewhere in the preface, in a discussion of the manuscript sources of the dictionary's entries, Somner mentions the Junius 11 manuscript of Old English verse on Biblical themes (Oxford, Bodleian

---

[3] As the title of his work suggests, Somner generally referred to the language he studied as 'Saxon'; the use of 'Old English' to refer to the vernacular language spoken in early mediaeval England would not be established until several centuries after Somner's time. In this article, I use 'Old English' throughout, in accordance with modern usage, except when directly quoting a source that uses a different term. I do not intend to imply by this approach that, for instance, Somner's understanding of the category 'Saxon' can be mapped directly onto the modern understanding of Old English; nor is my retention of terms such as '(Anglo-)Saxon' within quotations meant as an assertion of the appropriateness of such terms in a present-day context.
[4] Somner was not, however, the first lexicographer of Old English. For more detail on both Somner's dictionary and the unpublished works that preceded it, a useful resource is Hetherington (1980).
[5] 'plainly according to the various & varying dialect of the age or place'; this translation is taken from Somner's unpublished English version of the preface in Canterbury Cathedral Archives, CCA-DCc-ChAnt/M/352.

Library MS Junius 11, known to Somner as the 'Paraphrasis Saxonica'). Somner complains of the difficulty he had in translating it because of its language, which he found 'adeo veteri, obsoleto, poetico, tumido, affectato, mystico & ænigmatico [...] diu cogebar insistere' (Somner 1659, *Ad Lectorem* §7).[6]

Although Somner's dictionary gives us little sense of late Old English as a distinctively unusual sub-period, it does include vocabulary from some texts that would probably be described today as late or transitional in character. The status of these texts in Somner's work is somewhat ambiguous. Some of his comments imply that Old English came to an end with the Norman Conquest of 1066; s.v. *cniht*,[7] Somner discusses the word's semantic development from 'boy' to modern 'knight', concluding that 'in that notion [the sense 'knight'] I never find it used by the *English-Saxons*: after whose supplanting by the *Normans* it succeeded in the place of their ðegen, or *Thane.*'

Elsewhere in the dictionary, however, Somner makes use of texts that unambiguously postdate the Conquest. This is illustrated most strikingly by the entry for *domes-dæg*:

> Dies judicii. **the day of judgement.** *It.* Liber apud *Anglos* censualis: *Angliæ* Notitia. **the famous** *Domesday* **book in the Exchequer.** Monumentum sanè hoc non augustum minùs quàm antiquum, de quo utilis admodum & jucundus apud doctiss. *Spelmannum*, in optimo Glossario, tractatus.[8]

In this case, an entire sense (i.e. the part of the entry that begins 'Liber apud…') refers exclusively to the Domesday Book, a work that not only post-dates the Norman Conquest, but that could be said to count the Conquest as a driving factor in its production. While the entry does not directly mention the date of the Domesday Book, the fact that it post-dates the Conquest was well-known to Somner (and, presumably, to his intended

---

[6] 'old, obsolete, uncouth, poeticall, swelling, effected, mysticall [and] ænigmaticall …; & so full of strange hyperbata, & transpositions… I was enfourced to plod much'; translation Canterbury Cathedral Archives, CCA-DCc-ChAnt/M/352.

[7] Note that, wherever I cite a dictionary headword, I refer to the orthography and meaning recognized in that dictionary, regardless of whether it would be considered correct by modern scholarship.

[8] 'Judgement day. **the day of judgement.** *Also* a census book among the *English*: Record of *England*. **the famous** *Domesday* **book in the Exchequer.** Truly this monument is not less august than it is ancient, about which there is a very useful and pleasant discussion by the most learned *Spelman* in his very good Glossary.' [Translation my own.]

audience) and is made clear in the discussion by Henry Spelman (1626, s.v. *Domesdei*) to which Somner refers.

As a result, words such as *domes-dæg* occupy an ambiguous position in Somner's dictionary. The use of a post-Conquest source makes the status of the sense 'Domesday Book' as Old English questionable according to Somner's apparent policies as expressed elsewhere in the dictionary; nevertheless, its late, liminal status is not labeled or addressed directly. If, in Somner's conception of the history of English, there is a category equivalent to late Old English, it is at most only present by implication, and it seems likely that he gave little or no conscious thought to the matter.

3. Late Old English in eighteenth- and nineteenth-century dictionaries

In major dictionaries of Old English compiled after Somner's time, however, we see a growing consciousness of late Old English (or something broadly equivalent to it) as a distinctive period in the development of English. I present the following examples from eighteenth- and nineteenth-century dictionaries of Old English in chronological order of publication, as each of the dictionaries discussed in this section was substantially influenced by those that preceded it. Nevertheless, it will be seen that they display varying degrees of systematicity in their representations of late Old English thanks to the differing strategies and priorities of the lexicographers.

The idea of identifying lateness as a significant factor in the description of Old English lexis appears to have taken some time to establish itself among lexicographers of Old English. The revised and abridged version of Somner's dictionary attributed to Thomas Benson (1701) strips out Somner's preface and discursive comments, with the result that it presents almost all its vocabulary on a single level, with no distinction between standard and non-standard. The exception is a short appendix of just over 150 entries, entitled 'Voces Poeticæ cum interpretamentis, è Grammatica D. *Georgii Hickesii*, S.T.P.'[9] In this we can see an echo of Somner's comments on the difficult language of the Junius 11 poems; again, what gives a word marginal status is not its lateness, but its being confined to the register of poetry. However, we can also see here the introduction of an important new theme in the mention of the *Institutiones Grammaticae Anglo-Saxonicae et Moesogothicae* (1689) of George Hickes. Hickes's works, more specifically his subsequent *Linguarum vett. septentrionalium thesaurus grammatico-criticus et archæologicus* (1703-1705), would

[9] 'Poetic words with interpretations, from the Grammar of D. George Hickes, S.T.P.'

influence significantly the portrayal of marginal categories of Old English, and the emerging category of late Old English, in the next major dictionary of Old English, which appeared in the latter half of the eighteenth century.

3.1 Dictionarium Saxonico et Gothico-Latinum (1772)

Edward Lye's *Dictionarium Saxonico et Gothico-Latinum* appeared posthumously in 1772, having been completed and published by his friend Owen Manning. Although retaining much of Somner's material, it also introduced a considerable amount of new vocabulary to the Old English lexicographical tradition; important sources for these additions were the unpublished manuscripts of the seventeenth-century philologist Francis Junius the younger (Clunies Ross 1999; Clunies Ross and Collins 2004, 47-48). Significantly, Lye and Manning's work is the first dictionary of Old English of which I am aware that shows signs of using a labeling system to identify a distinct category of late vocabulary. Such a system would not appear again in the lexicography of Old English until the end of the nineteenth century, and not in a full-scale dictionary of Old English (as opposed to an abridged or student's dictionary) until the *Dictionary of Old English* in the mid-twentieth century.

The most common label used in Lye and Manning's dictionary is 'Norm[an] Sax[on]', though—with fewer than thirty occurrences—its impact on the dictionary as a whole is still minimal. The significance of this label is not discussed at any point in the dictionary, and therefore must be deduced by users from context and their general knowledge. It is clear that the act of labeling these entries establishes them as a marked category, and the term chosen as a label brings to mind two principles that may be being used to establish this markedness. The first of these is language contact, with *Norman* specifying that we are dealing with French influences on the English lexicon introduced after the Norman Conquest of 1066. As has been seen above, however, the historical event of the Conquest was already established in Old English lexicography (as in wider usage) as a possible dividing point between Old English proper and a more nebulous category that might be identified as late Old English. On this basis, it is possible to interpret the 'Norm[an] Sax[on]' label as reflecting lateness as a general principle as well as French influence.

This interpretation is strengthened by an examination of the entries labeled in this way. Some words so labeled are distinctly French in form, such as *pais* (peace). Others are represented as corrupt spellings (implied to be by Anglo-Norman scribes) of standard Old English forms, as illustrated by comments such as 'sei. *Norm. Sax.* pro Sægan. Dicere,' (to say), or expressed more clearly s.v. *te*: 'occurrit, apud *Norm. Sax.* scriptores, pro Se articulo...'[10] Some words, however, receive the 'Norm[an] Sax[on]' label even though they are neither obviously French in form nor explicitly presented as variants of a more correct, non-French-influenced, standard; examples are *lið-mon* (sailor) and *ploge* (plow, here strictly a plow team). If French influence cannot be readily identified as the reason for the markedness of these entries, we get the sense that these labels are, at least to a certain extent, capturing a category of late Old English.

Two other labels, 'Anglo-Norm[an]' and 'Semi-Sax[on]', appear in the usage of the Lye-Manning dictionary to be approximately synonymous with 'Norm[an] Sax[on]', though they are used much less frequently. A small number of entries are also given the label 'Dano-Sax[on]', although this is a term less clearly correlated with lateness than the others.[11] However, as all of these labels are considerably less frequent than 'Norm[an] Sax[on]', their overall contribution to the dictionary's representation of late Old English is less significant.

If we examine the 'Norman Saxon' entries in more detail, some interesting patterns emerge. The majority are drawn from a small group of texts, of which the most commonly cited is the Peterborough Chronicle, also known as text E of the Anglo-Saxon Chronicle, preserved in Oxford, Bodleian Library MS Laud Misc. 636.[12] This is appropriate, given that the two continuations of the Peterborough Chronicle (consisting of annals treating the years 1122-1131 and 1132-1154) are still viewed by modern scholarship as key witnesses to the transition between Old and Middle English. From the evidence of these citations, Lye clearly had some sense of this, but it is worth noting that some 'Norman Saxon' words are in fact drawn from entries that predate the continuations; for instance, *lit* (ship), from the entry for 1052, and *snacca* (small, fast ship), from the entry for 1066.

The majority of the remaining cited texts are easily verified as being taken from George Hickes' *Thesaurus* (1703-1705), which includes several samples of late texts. However, even when the *Thesaurus* extracts lists of characteristic vocabulary from these texts, only occasional items from them

---

[10] 'It occurs among Norman Saxon writers for the article Se...'

[11] Even more than is the case for 'Norman Saxon', it is hard to map 'Dano-Saxon' meaningfully onto modern categories or ideas of language change; for a fuller discussion of the implications of 'Dano-Saxon' as a category, see Cain (2010, 734-737), who analyzes the use of this and related terms in George Hickes' *Thesaurus*.

[12] Lye did not work directly from the manuscript, but rather, as is apparent from the page numbers included in some citations, from the 1692 edition of the Chronicle by Edmund Gibson.

are transferred into Lye-Manning. The overall impression of the Lye-Manning *Dictionarium* is that of a lexicographer without any particular interest in identifying transitional elements of Old English vocabulary; the labeling of entries as Norman Saxon and similar is not thorough, and is heavily dependent on a small number of texts that had already been categorized in this way by previous scholars.

The use of the 'Norm[an] Sax[on]' label in the Lye-Manning dictionary seems at first glance to establish a clear category of late Old English, a clear way of labeling it within the dictionary and a clear way of understanding it—in terms of post-Conquest contact between English and French. In practice, however, the small number of labeled entries and the inconsistent use of Hickes's classifications in determining the application of these labels mean that it is hard to see Lye and Manning's dictionary as showing a fully-developed understanding of late Old English as a distinct category.

3.2 A Dictionary of the Anglo-Saxon Language (1838)

Joseph Bosworth's *A Dictionary of the Anglo-Saxon Language,* published in 1838, is even more complex in its representation of late Old English than Lye and Manning's dictionary. This complexity does not arise from the dictionary's entries; these are frequently taken verbatim from Lye and Manning, but the labeling system used in the older dictionary has been discarded, meaning that Norman Saxon words (and other categorizations used by Lye and Manning) are no longer marked. In addition to discarding the old, Bosworth does not introduce any new labels.[13]

Bosworth's innovation in the representation of late Old English is found instead in his dictionary's lengthy preface. This preface is a feature not found in other, similar dictionaries of Old English. It is more than 170 pages long, and the majority of its length is dedicated not to describing the features and methodology of the dictionary, or even to giving a general linguistic account of Old English, but to discussing all the major Germanic languages, their development, the relationships between them and the history of their speakers. In this respect, Bosworth's dictionary can be compared to contemporary synchronic dictionaries of English, in which it was already a well-established convention to include histories of English (sometimes reaching back to the origins of language) as part of an encyclopedic front matter (Rodríguez-Álvarez 2009).

[13] Useful background to Bosworth's 1838 dictionary and his use of material from Lye and Manning is to be found in Baker (2003, 287-291).

Bosworth's preface is a semi-independent piece of scholarship from the dictionary proper, and was indeed later published as a separate volume (Bosworth 1848). The claims it makes about the history and periodization of English are not necessarily reflected in Bosworth's lexicographical methodology, which—as I have already noted—was heavily shaped by his use of Lye-Manning. As a result, the dictionary's overall portrayal of late Old English is complex and self-contradictory. Nevertheless, the wide historical and contextual overview of the preface offers the clearest picture we have yet seen of the relationship between Old English and other periods or varieties of English, and contributes to the dictionary's more detailed picture of late Old English as a distinctive category.

Although Bosworth does not use Lye and Manning's labels in his dictionary entries, there is a certain element of continuity in the sense that the term 'Semi-Saxon', used extremely infrequently as a label by Lye and Manning, is used by Bosworth in his preface to describe a group of texts that might be identified as late Old English. The applications and implications of the term 'Semi-Saxon' in general historical and linguistic scholarship are discussed by Matthews (1999, xxvii-xxxii) in the context of the nineteenth-century emergence of Middle English as a distinct discipline. Matthews demonstrates that the term, which had already appeared in the work of Hickes, Lye and Manning and others, was revived in the 1830s in works such as Benjamin Thorpe's *Analecta Anglo-Saxonica* (1834), as one of several labels used to divide up the—by now complex and contested—territory of late Old English as a transitional period between Old and Middle.

Bosworth's preface reflects these contemporary interests, using 'Semi-Saxon' to identify a category of late Old English texts, of which he writes that 'a few extracts are only given, that they may be compared with the specimen of pure Anglo-Saxon' (Bosworth 1838, xxii). In other words, Bosworth clearly seems to be using 'Semi-Saxon' to establish late Old English as a subordinate category. The texts to which he applies this label are *The Grave* (Bosworth 1838, xxiv), the *Ormulum* (Bosworth 1838, xxiv-xxv) and the *Ayenbite of Inwit* (Bosworth 1838, xxvi). However, Bosworth's perception of lateness as a significant feature of medieval English texts appears to extend beyond those he labels as Semi-Saxon. For instance, his 'extracts' also include passages from the second continuation of the Peterborough Chronicle (Bosworth 1838, xxiii), Robert of Gloucester's rhyming chronicle (Bosworth 1838, xxv) and Wycliffe's Bible (Bosworth 1838, xxv-xxvi), which he does not identify as examples of Semi-Saxon. Furthermore, in the dictionary entries, I have only found a single citation from one of the three texts that are called Semi-Saxon in the preface: the *Ormulum*, s.v. *swaran*, an entry inherited from Lye-Manning. On the whole, Bosworth seems to have considered so-called Semi-Saxon texts as sufficiently different from Old English proper to be excluded from his citations, even if they found a small place in his discursive preface.

This does not mean, however, that there is no overlap between Bosworth's preface and his entries with respect to the portrayal of lateness as a distinctive category within Old English. Elsewhere in the introduction, when Bosworth is actually describing the editorial principles of his dictionary, he clearly implies the presence of another category of late Old English texts. The following passage is taken from Bosworth's description of the chronological scope of his dictionary:

> It was originally intended to exclude all impure Anglo-Saxon words, and to introduce none of a later date than AD 1100. Subsequently it was found desirable to take a wider range, and to include some terms of a more recent formation. These are mostly from the Anglo-Saxon Chronicle, with their date affixed. As the authors are always quoted, the age and purity of a word can at once be seen. (Bosworth 1838, clxxii)

In this we see a rather grudging partial inclusion of vocabulary from post-1100 texts, which Bosworth originally wished to exclude specifically because of its lateness. This group of texts is different from the Semi-Saxon category described elsewhere in the introduction, because it contains different texts, specifically the later entries in the Anglo-Saxon Chronicle. (As noted above, the second continuation of the Chronicle's Peterborough recension was included in Bosworth's 'extracts' but not described as Semi-Saxon.) It is also distinct from Bosworth's conception of Semi-Saxon because it is covered, at least partly, in the dictionary's entries. However, Bosworth's treatment of this material still retains the idea—arguably even more strongly than in the other dictionaries discussed here—of late Old English as a corruption of an earlier standard. This passage also serves to demonstrate once again the interaction between general, theoretical concerns and specific, practical problems of lexicography in the creation of the category of late Old English. Bosworth's creation of a marked category of late vocabulary in his dictionary entries (by noting the author or source text, and the entry date in the case of material from the Anglo-Saxon Chronicle) is revealed to be a compromise solution arising from his inability to define a precise date for the end of the Old English period.

3.3 An Anglo-Saxon Dictionary (1882-98)

Bosworth's posthumously-published *An Anglo-Saxon Dictionary* (1882-98), which was completed by T. Northcote Toller, offers its users a very different picture of lateness in Old English to that seen in his 1838 dictionary.[14] Unlike the earlier dictionary, *An Anglo-Saxon Dictionary* does not have a detailed introduction. It also lacks systematic entry labeling to indicate lateness. Nevertheless, comments in a few individual entries bear witness to the developing idea of late Old English as a category to explain deviations from the standard form of the language.[15]

Thus, s.v. *steallere*, we find the following observation given in square brackets after the definition ('a marshall'): 'The word occurs only in late documents; the passages given belong to Edward the Confessor's reign.' An even clearer expression of the notion that late Old English is a corruption of the standard can be found s.v. *híd* (hide of land) in Toller's supplement to the dictionary (1921): 'The nominatives *híde, gyrde* […] are perhaps really incorrect late forms.' Interestingly, most of the comments of this type that mention lateness as a distinctive feature are clustered in either the second half of the main dictionary or in Toller's 1921 supplement. This means that they do not appear in the part of the dictionary that is based most closely on Bosworth's work, but in the substantial additional material contributed by Toller after Bosworth's death.

In the Preliminary Notice to Parts I and II of the dictionary, published in 1882, Toller informs his readers that 'at the time of his [Bosworth's] death, only the 288 pages which form Part I of the present issue had been finally revised by him. So much progress, however, had been made with some succeeding sheets that it would have been a matter of considerable difficulty to make any but slight alterations in them'. Thus, the entries *A-firgen* are primarily Bosworth's work, only lightly edited by Toller (Bankert 2003, 304). Progressing through the alphabet, more and more of the material is Toller's own and, as Peter Baker demonstrates (2003, 294-297), it is possible to track the development of Toller's methodology as he gradually moved away from the model established by Bosworth and introduced his own innovations, entry style and inclusion policies in the process of bringing the dictionary to completion.

---

[14] For a more detailed account of the dictionary and the extent of Toller's contribution to it, see Baker (2003, 291-300) and Bankert (2003).

[15] My identification of these entries was heavily dependent on the digitized version of both Bosworth and Toller's dictionary and Toller's supplement, currently maintained by Ondřej Tichý at Charles University in Prague and freely available at www.bosworthtoller.com.

I have identified only two entries that directly mention lateness in the range A-F of the main dictionary (the part that is mostly Bosworth's work). S.v. *burh* is the comment, with reference to a quoted charter, that 'The style of the Anglo-Saxon indicates a late date, perhaps about 1100 or 1200'. The entry for *blóma* (lump of metal) employs the older term 'Semi-Saxon'—discussed above in this article but not otherwise present in Bosworth and Toller's dictionary—noting that further evidence for the word can be found 'in a Semi-Saxon glossary of the 12th century'. All other entries in the dictionary in which lateness is mentioned as a linguistically distinctive feature are found either in the range H-Z (*hird, land-scearu, mál*, *steallere, stig-weard, swát, swólig* and *underling*) or else in Toller's 1921 supplement to the dictionary (*á-hlocian, ge-myþe, hærfest, híd, hwá, níþ-hell* and *norþ*).

From the evidence of the Bosworth-Toller dictionary as it was published, Bosworth had planned something rather similar to his 1838 dictionary in the sense of not using labels or comments within entries to establish late Old English as a distinct category. Toller's entry-writing, although using ad hoc comments rather than a consistent labeling system, makes much clearer the concept of late Old English as a recognizable sub-period. Nevertheless, as there are few such comments on lateness in either the dictionary or Toller's supplement—and even fewer in the early entries, on which Toller's influence was minimal—it is hard to describe the recognition of late Old English as a consistent feature of the dictionary as a whole.

It is worth noting that many of the comments on lateness found in Bosworth and Toller's dictionary have a scope smaller than an entire entry. They are attached to an illustrative quotation within the entry and refer not to a word but to a particular token of that word. This reinforces the sense that Toller's concept of late Old English is invoked on a case-by-case basis to explain inconsistencies, rather than functioning as a category in its own right.

The application of comments to individual quotations also allows us to associate Toller's 'late' descriptor with particular texts, of which (allowing for the small total number of entries containing comments) there is a reasonable variety. Texts implied by Toller's comments to be typical of late Old English include not only eleventh- and twelfth-century entries from the Anglo-Saxon Chronicle, but also various charters, the Bodleian Hatton 38 manuscript of the West-Saxon gospels, the Worcester manuscript of Ælfric's Grammar and Glossary, the life of St Margaret in Corpus Christi College Cambridge MS 303 and a few others. This might be seen as the beginnings of a move away from the pattern established by Lye and Manning—and continued in Bosworth's *A Dictionary of the Anglo-Saxon*

Language—in which lateness was associated primarily with a small group of named literary texts.

It is worth noting that more than half of the entries containing comments on lateness in Bosworth and Toller's dictionary have an equivalent in Bosworth's earlier dictionary, although I have found no instances in which the earlier dictionary entry also includes an equivalent comment. This is consistent with the general characterization of entries in Bosworth's dictionary given above.

However, given that the earlier eighteenth-century dictionary of Lye and Manning includes a partial labeling system that expresses lateness, it is interesting that there is no overlap between entries containing a comment on lateness in Bosworth and Toller's dictionary (including Toller's supplement) and those labeled as Norman Saxon (or similar) by Lye and Manning. This is partly a reflection of the fact that, although the two dictionaries have created categories that are functionally similar, they have different ideas about what words and texts belong in those categories. For example, a significant number of words labeled as Norman Saxon by Lye and Manning are French loanwords. This is consistent both with the choice of 'Norman Saxon' as a label and with Lye and Manning's general understanding of the idea of lateness in terms of contact with French. However, obviously French loanwords are usually excluded from Bosworth and Toller's dictionary, which appears to have stricter inclusion criteria when it comes to the treatment of borrowed vocabulary. Nevertheless, this explanation cannot fully account for the lack of overlap in the two dictionaries. Although there is no overlap on the word level, some *texts*, such as the eleventh- and twelfth-century entries in the Peterborough Chronicle, appear both in Lye-Manning as 'Norman-Saxon' and in Bosworth-Toller as 'late'. This inconsistency reinforces the impression that both dictionaries represent late Old English fairly unsystematically, rather than as a well-defined period.

3.4 A Concise Anglo-Saxon Dictionary (1894)

J.R. Clark Hall's *Concise Anglo-Saxon Dictionary* was first published in 1894; the most recent edition is the fourth, which includes a supplement by Herbert D. Meritt (Hall & Meritt 1960). It was originally intended as a concise dictionary aimed at students and other non-specialist users (Hall 1894, iv). Despite this, Hall wished to make his dictionary 'as comprehensive as possible' (Hall 1894, iii), and the resulting work has enjoyed long-lasting popularity and authority; for instance, the *Thesaurus of Old English* is largely dependent on the coverage and definitions of the dictionary's fourth edition (Roberts & Kay 2017).

Nevertheless, reflecting its original purpose, the *Concise Anglo-Saxon Dictionary* tends towards a user-friendly simplicity and economy of space that leaves far fewer opportunities for a discussion of lateness. Labeling is relatively minimal, and although 'LWS' does occasionally appear as a marker of Late West Saxon forms, this term is used in opposition to the Early West Saxon dialect that the dictionary adopts as a standard, and it does not appear to have any bearing on lateness as considered with respect to Old English as a whole.

More informative are the references, first added in the second edition, to A. Napier's "Contributions to Old English Lexicography" (1906). In the list of abbreviations, under 'NC', it is observed that these are 'mostly late texts' (Hall & Meritt 1960, xiii). Referring back to Napier's work, we find that 'the large majority [of words in the "Contributions"] are taken from MSS. of the eleventh century, but I have also included words from a few twelfth-century MSS., as in many cases these are copies of eleventh-century originals, and in most cases were probably in use before 1100' (Napier 1906, 265-266). However, this detail is lost in the dictionary; since 'NC' denotes a (secondary) source rather than an inherent property of the medieval text, not all of the entries that might be considered late are labeled 'NC' and not all entries labeled in this way are necessarily late. As a result, the label is in many ways similar to the 'Norman Saxon' labeling seen in Lye and Manning's dictionary, and can hardly be said to establish lateness as a distinctive linguistic category.

Similarly, the third edition of Hall's dictionary introduced 'a considerable number of words from twelfth-century texts, which have not been recorded in B[osworth] T[oller]… as there seemed to the writer that there was no sufficient ground for their exclusion' (Hall & Meritt 1960, vii). This change, and its announcement in the preface, might seem to imply a shift in the treatment of lateness, but as there is no further discussion of the twelfth-century texts in question, it is hard to draw any firm conclusions from this about late Old English as a category in the dictionary. Only one text, the Eadwine Psalter, is conclusively identified as 'late' in the list of sources (Hall & Meritt 1960, xi), while the homilies in Oxford, Bodleian Library MS Bodley 343 and London, British Library MS Cotton Vespasian D. xiv are noted to be '12th century' (Hall & Meritt 1960, x, xiv); given how isolated these references are, it is again hard to extrapolate from this to a broader category of lateness. Despite its importance in other respects, the *Concise Anglo-Saxon Dictionary* does not, it seems, make any notable advances in the representation of lateness.

3.5 The Student's Dictionary of Anglo-Saxon (1896)

However, another concise dictionary for students, published not long after the *Concise Anglo-Saxon Dictionary*, provides considerably more grounds for discussion of lateness. This is Henry Sweet's *The Student's Dictionary of Anglo-Saxon*, first published in 1896. Although shorter even than Hall's dictionary, it represents some important developments in the treatment of late Old English vocabulary.

Significantly for the developing understanding of late Old English, it was Sweet who formulated what is now the classic tripartite periodization of the history of English, in which 'O[ld] E[nglish] may be defined as the period of *full* endings *(mōna, sunne, sunu, stānas*), M[iddle] E[nglish] as the period of *levelled* endings (*mōne, sunne, sune, stǭnes),* M[oder]n E[nglish] as the period of *lost* endings (*moon, sun, son, stones* = stounz)' (Sweet 1892, 211, §594). Sweet goes on to describe numerous subdivisions of the three main periods, including 'Late Old English' (900-1100), 'Transition Old English' (1100-1200) and 'Early Middle English' (1200-1300). However, the point most significant for the current discussion is that Sweet's influence in this area makes the approach to lateness he took in his student's dictionary of Old English especially interesting.

It is typical of both the requirements of a student dictionary and of Sweet's style that the preface to *The Student's Dictionary of Anglo-Saxon* is detailed in setting out Sweet's lexicographical principles (and in criticizing those of previous lexicographers). One of these principles is the treatment of late vocabulary, to which Sweet dedicates a short section (1896, vii-viii). This singling out of lateness as a distinct topic requiring specific discussion is a feature not seen to the same extent in earlier dictionaries, even in Bosworth's lengthy preface.

In Sweet's preface, we see the continuation of several themes already identified. Late Old English is presented as a problematic intrusion which Sweet (like Bosworth before him) would exclude from his dictionary if he could; the presence of late Old English vocabulary in earlier publications is said to be the fault of 'uncritical compilers' (Sweet 1896, vii). Those elements of late Old English that Sweet is prepared to include in his dictionary are (again echoing Bosworth's comments before him) from the Anglo-Saxon Chronicle, 'partly on the chance of their being really older than their first occurrence, partly because of the continuity and great importance of the Chronicle' (Sweet 1896, vii).

However, Sweet also introduced an important innovation to which I have already alluded: entries for words deemed late are labeled, on a much larger scale than the infrequent and inconsistent labeling seen in Lye and Manning's dictionary. Sweet's labels are also more clearly associated (solely) with lateness rather than covering other factors such as dialectal variation and language contact, and their intended meanings are explained in the dictionary's front matter. Sweet recognizes two degrees of lateness; in addition to 'L.' for 'late', he uses 'vL.' for 'very late'. This appears to be more than a subjective distinction, as a specific criterion is given for the application of the latter label; 'very late' words 'first occur towards the end of the eleventh century' (Sweet 1896, viii). As we will see, the recognition of two degrees of lateness is a strategy also used by the most important Old English dictionary of the present day, the *Dictionary of Old English.*

4. Current approaches: the *Dictionary of Old English*

The *Dictionary of Old English* project, based at the University of Toronto, represents a break from the Old English lexicographical tradition in a number of ways. Although it will, when complete, meet the same linguistic needs as earlier dictionaries of Old English, it also offers new ways of thinking about late Old English as a distinct sub-period.[16] The *Dictionary of Old English* is able to draw on modern scholarship in the field that was not available to earlier lexicographers, such as the increasing number of textual editions.[17] In addition, however, I argue that the novel methodologies and format of the dictionary itself enable the category of late Old English to be understood differently.

The *Dictionary of Old English* was intended from its earliest stages to re-examine the vocabulary of Old English rather than being dependent on existing dictionaries. Its entries are based on a purpose-built corpus (Healey et al. 2009), producing what was envisaged as 'a new dictionary afresh from the texts' (Leyerle 1971, 279). The existence of a corpus cannot, of course, guarantee an objective treatment of lateness in Old English; although the *Dictionary of Old English* corpus aims to include at least one copy of all known Old English texts, this does not resolve the deeper issue of which texts (especially late texts) count as Old English and therefore belong in the corpus. Still, it is important to note that the creation of the *Dictionary of Old*

[16] For more background on the *Dictionary of Old English* and its origins, Adams (2009, 345-351) provides a useful starting point.
[17] Texts that are treated as late in the *Dictionary of Old English* (by labeling as discussed below) and which were not readily available to earlier lexicographers include, among many others, the Old English *Life of St Nicholas*, for which the *Dictionary of Old English* relies on Treharne's edition (1997), and the Worcester manuscript of Ælfric's *Grammar* and *Glossary*, for which Butler's edition (1981) is used, superseding a partial edition (Phillipps 1838) made in the mid-nineteenth century and cited by Toller in the latter half of the Bosworth-Toller dictionary.

English corpus effectively increased the amount of late Old English available to researchers, since some texts—which, according to the criteria used by the dictionary, are late Old English—were edited especially for inclusion in the corpus. An examination of the *DOE'*s 'List of Texts' mentions the existence of a 'DOE [or 'Dictionary of Old English'] transcript' for more than 100 Old English texts (mostly charters).

While lexicographers such as Bosworth and Sweet seem to have wished to identify late Old English primarily to enable them to exclude as much of it from their dictionaries as was feasible, the *Dictionary of Old English* takes an altogether different approach, with inclusiveness as the primary goal. As one of the editors wrote, 'we are trying to ensure that no material of the transitional period between Old and Middle English is overlooked, and occasionally we have extended the upper limit of our period in order to encompass late twelfth- and thirteenth-century texts which show the development of the earlier tradition' (Healey 1997, 57).

Another feature of the *Dictionary of Old English* that has an important impact on the representation of late Old English is the decision to list attested spellings of dictionary headwords. Although earlier dictionaries of Old English have often noted especially frequent or unusual spelling variants, the *Dictionary of Old English* provides much more information, recording all spellings found in the *Dictionary of Old English* corpus and many more from manuscripts not represented in the corpus. In terms of the representation of late Old English, what is significant about this is that the attested spellings are labeled to indicate two different degrees of lateness. Spellings found only in twelfth-century manuscripts are marked 'xii'; spellings found in manuscripts of the thirteenth century or later are not only marked with Roman numerals but are also placed in a separate section of the attested spellings field, under the heading 'late'.

The recognition of two degrees of lateness through consistent labeling is reminiscent of Sweet's approach, although the *Dictionary of Old English* interprets the category of lateness in a very different way; Sweet's 'very late' begins in the late eleventh century, while the 'late' label used in the *Dictionary of Old English* only starts to be applied more than a hundred years later, at the beginning of the thirteenth century. Once again, we see that agreement in how the concept of lateness is represented lexicographically is to a certain degree independent of agreement in the criteria by which a word's inclusion in the late category is decided.

It is important to bear in mind here that the *Dictionary of Old English* establishes the category of late Old English on a different level from most other Old English dictionaries, because its labels are attached to particular orthographic forms, and thus to particular manuscripts, rather than to the word as a whole. Gathering such a volume of information on spelling variants is labor and resource intensive, and displaying it requires more space in the dictionary entry than would be possible in most print dictionaries. The way the *Dictionary of Old English* represents late Old English through variant spellings is hence dependent on the technology underlying the dictionary as a whole: the corpus of texts (and ready access to editions) that allows the data to be gathered efficiently, and the web publication that removes many of the space constraints of earlier, print dictionaries.[18]

Although Bosworth and Toller's dictionary was not so ambitious as to attempt the extensive documentation of variant spellings, I have already observed that its comments on lateness are relatively frequently applied to particular texts or manuscripts rather than to entire entries. In this respect, it could be seen as a precursor of the fine-grained approach to lateness that is created in the *Dictionary of Old English* by the labeling of variant spellings. However, the *Dictionary of Old English* achieves this through systematically applied labels rather than sporadic comments, giving a better sense of late Old English as a complete (if subordinate) system, rather than a set of ad hoc exceptions.

5. Conclusions

There is a general trend in which a category of late vocabulary plays an increasingly significant role in the lexicography of Old English; lateness is an important and well signposted issue in the present-day *Dictionary of Old English*, but seems to have been barely considered by Somner in the seventeenth-century *Dictionarium Saxonico-Latino-Anglicum*. It seems likely that this difference can be explained in large part by the development of more rigid systems of periodization. For instance, Sweet, who developed a complex and fine-grained account of the successive periods of development of the English language, chose in his dictionary to use an innovatively precise labeling system that would better indicate (and justify) how so-called late words fitted into his tightly defined category of Old English. Somner, in contrast, seems content to sketch the boundaries of the Old English period in much broader strokes, and hence, it might be argued, the idea of lateness becomes less important and does not need to be signaled as clearly within the dictionary.

[18] Before being published on the web, the *Dictionary of Old English* was also available on microfiche and CD-ROM, but these, too, are formats that make the publication of lengthy entries more feasible.

Another, related, explanation for the general trend concerns the amount of data available to lexicographers; as Old English studies progressed, more texts became available to scholars, and these texts could be dated with more accuracy. I suggest that the emergence of a clearer late Old English category within dictionaries arose in part from the ability to make more fine-grained distinctions of this kind.

However, the emergence of a late Old English category was not simply a case of continuous refinement. As my chronological account has made clear, not all lexicographers of Old English necessarily provided their users with more detail regarding lateness than had their predecessors. Some informative elements in earlier dictionaries were abandoned by subsequent lexicographers who in other respects were hewing closely to the model offered by the older work. This is the case for Bosworth's lengthy preface, which—at least in terms of Old English lexicography—was a developmental dead end not adopted even by the Bosworth-Toller dictionary published later in the same century. In other cases, lexicographical techniques that could establish late Old English as a distinct category came in and out of existence; this could be said to be the case for the various experimentations with labeling seen in the dictionaries of Lye and Manning, Sweet and the *Dictionary of Old English.* As a result, different dictionaries of Old English may offer more clarity about lateness in Old English in one respect, but less in others; Bosworth's dictionary, with its preface, offers more historical context for the concept of lateness, but is followed by Bosworth and Toller's dictionary, which omits this contextualization but is significant in how it begins to use the idea of lateness to give much more specific and targeted discussions of individual manuscript readings.

These varying approaches can perhaps be explained, at least in part, by the observation that the field of Old English lexicography was always somewhat specialized; although this article does not cover all dictionaries of Old English published within the last five centuries, it does address most of the important general-purpose scholarly dictionaries. Within this field, the concept of lateness is one minor concern among many. Arguably, there were simply never enough lexicographers working on the topic for strong traditions or conventions to develop, leading to experimental and innovative approaches that were not necessarily retained. Furthermore, most of the lexicographers discussed in this article had distinctly different criteria for what counted as late Old English, even if they shared a conviction that lateness was worth marking and discussing. If they had different assumptions about what late Old English was, it is perhaps not surprising that they also chose to represent it in different ways.

Dictionaries of Old English vary in the extent to which late Old English is portrayed as a distinct sub-period, and when the concept is used, it may be indicated in a variety of ways. However, in almost all the dictionaries that identify a late category, we find the repeated idea that late Old English is not only subordinate to the standard form of the language, but a corruption of it. This is seen most clearly in the dictionaries of Bosworth, Toller and Sweet, all of whom express their reservations about the 'purity' and reliability of sources they consider late. This rather dismissive approach only receives a serious challenge in the *Dictionary of Old English*, which uses an approach that is generally more theory- and value-neutral, while still relegating late material firmly to subordinate status at the edges of the entry.

The examples of Old English dictionaries presented in this article illustrate how historical lexicographers, variously influenced by each other, by external scholarship and by the conventions and limitations of dictionary form, have gradually approached the situation seen today, in which it is accepted that late Old English should be addressed as a significant issue in an Old English dictionary. Although late Old English may remain, in relative terms, a minor issue to these lexicographers, by examining its portrayal in more detail we have seen the variety of techniques that are used to bring it to the attention of readers—often while simultaneously downplaying its wider historical significance by presenting it as merely a convenient label to explain away apparent corruption or error. It has also demonstrated some of the ways we can find hints in dictionaries about a period in the history of English that often seems to have been treated as unimportant almost by definition.

Bibliography

Adams, Michael. 2009. "The Period Dictionaries". In *The Oxford History of English Lexicography*, Volume I, edited by A.P. Cowie, 326-52. Oxford: Clarendon Press.

Bailey, Richard. 1990. "The Period Dictionary III: English". In *Wörterbücher / Dictionaries / Dictionnaires: Ein internationales Handbuch zur Lexikographie / An International Encyclopedia of Lexicography / Encyclopédie internationale de lexicographie, 2 Teilband,* edited by Rufus Gouws, Ulrich Heid, Wolfgand Schweickard and Herbert Ernst Wiegand, 1436-57. Berlin: De Gruyter Mouton.

Baker, Peter S. 2003. "Toller at school: Joseph Bosworth, T. Northcote Toller and the progress of Old English lexicography in the nineteenth century". In *Textual and Material Culture in Anglo-Saxon England:*

Thomas Northcote Toller and the Toller Memorial Lectures, Publications of the Manchester Centre for Anglo-Saxon Studies 1, edited by Donald Scragg, 283-300. Cambridge: D.S. Brewer.

Bankert, Dabney Anderson. 2003. "T. Northcote Toller and the Making of the *Supplement* to the *Anglo-Saxon Dictionary*". In *Textual and Material Culture in Anglo-Saxon England: Thomas Northcote Toller and the Toller Memorial Lectures*, Publications of the Manchester Centre for Anglo-Saxon Studies 1, edited by Donald Scragg, 301-22. Cambridge: D.S. Brewer.

Benson, Thomas. 1701. *Vocabularium Anglo-Saxonicum, lexico Gul. Somneri magna parte auctius.* Oxford: Sheldonian Theatre.

Bosworth, Joseph. 1838. *A Dictionary of the Anglo-Saxon Language: Containing the Accentuation—the Grammatical Inflections—the Irregular Words Referred to Their Themes—the Parallel Terms, from the Other Gothic Languages—the Meaning of the Anglo-Saxon in English and Latin—and Copious English and Latin Indexes, Serving as a Dictionary of English and Anglo-Saxon, as Well as of Latin and Anglo-Saxon*. London: Longman, Rees, Orme, Brown, Green, and Longman.

Bosworth, Joseph. 1848. *The Origin of the English, Germanic and Scandinavian Languages, and Nations.* London: Longman, Brown, Green, and Longmans.

Bosworth, Joseph & Toller, T. Northcote. 1882-98. *An Anglo-Saxon Dictionary*. Oxford: Clarendon Press.

Butler, Marilyn S. 1981. *An Edition of the Early Middle English Copy of Ælfric's "Grammar" and "Glossary" in Worcester Cathedral MS F 174.* PhD diss., Pennsylvania State University.

Cain, Christopher M. 2010. "George Hickes and the 'Invention' of the Old English Dialects". *The Review of English Studies,* New Series, vol. 61, no. 252: 729-48.

Cameron, Angus, Ashley Crandell Amos, Antonette diPaolo Healey et. al. 2018. *Dictionary of Old English: A to I online.* Toronto: Dictionary of Old English Project. Accessed January 7, 2020. http://tapor.library.utoronto.ca/doe.

Clunies Ross, Margaret. 1999. "Revaluing the work of Edward Lye, an eighteenth-century septentrional scholar". *Studies in Medievalism* 9: 66-79.

Clunies Ross, Margaret & Amanda J. Collins. 2004. *The Correspondence of Edward Lye*. Publications of the Dictionary of Old English 6. Toronto: Pontifical Institute of Medieval Studies.

Coleridge, Herbert. 1859. *A Glossarial Index to the Printed English Literature of the Thirteenth Century.* London: Trübner & Co.

Gibson, Edmund. 1692. *Chronicon Saxonicum.* Oxford: Sheldonian Theatre.

Hall, John R. Clark. 1894. *A Concise Anglo-Saxon Dictionary for the Use of Students*. London: Swan Sonnenschein & Co.

Hall, John R. Clark & Herbert D. Meritt. 1960. *A Concise Anglo-Saxon Dictionary.* Fourth edition. Cambridge: Cambridge University Press. Reprinted as: Medieval Academy Reprints for Teaching no. 14. Toronto: University of Toronto Press.

Healey, Antonette diPaolo. 1997. "Dictionary of Old English". In *Dictionaries of Medieval Germanic Languages: A Survey of Current Lexicographical Projects,* edited by K.H. van Dalen-Oskam, K.A.C. Depuydt, W.J.J. Pijnenburg and T.H. Schoonheim, 55-62. Turnhout: Brepols.

Healey, Antonette diPaolo, with John Price Wilkin and Xin Xiang. 2009. *Dictionary of Old English Web Corpus.* Toronto: Dictionary of Old English Project. Accessed January 7, 2020. http://tapor.library.utoronto.ca/doecorpus.

Hetherington, M. Sue. 1980. *The Beginnings of Old English Lexicography*. Spicewood, TX: published by the author.

Hickes, George. 1689. *Institutiones grammaticae Anglo-Saxonicae et Moesogothicae*. Oxford: Sheldonian Theatre.

Hickes, George. 1703-5. *Linguarum vett. septentrionalium thesaurus grammatico-criticus et archaeologicus.* Oxford: Sheldonian Theatre.

Kitson, Peter R. 1997. "When did Middle English begin? Later than you think!" In *Studies in Middle English Linguistics*, edited by Jacek Fisiak, 221-69. Berlin/New York: Mouton de Gruyter.

Lee, Stuart. 2000. "Oxford, Bodleian Library, MS Laud Misc 381: William L'Isle, Ælfric, and the *Ancrene Wisse*". In *The Recovery of Old English: Anglo-Saxon Studies in the Sixteenth and Seventeenth Centuries,* edited by Timothy Graham. 207-47. Kalamazoo: Medieval Institute Publications.

Leyerle, John. 1971. "'The Dictionary of Old English': A progress report". *Computers and the Humanities* 5 (5): 279-283.

Lutz, Angelika. 2002. "When did English begin?" In *Sounds, Words, Texts and Change: Selected Papers from 11 ICEHL, Santiago de Compostela, 7-11 September 2000*, edited by Teresa Fanego, Belén Méndez-Naya and Elena Seoane, 145-70. Amsterdam/Philadelphia: John Benjamins.

Malone, Kemp. 1930. "When did Middle English begin?" In *Curme Volume of Linguistic Studies*, edited by James Taft Hatfield, Werner Leopold and A.J. Friedrich Zieglschmid, 110-17. Baltimore: Waverley.

Matthews, David. 1999. *The Making of Middle English, 1765-1910*. Medieval Cultures 18. Minneapolis, London: University of Minnesota Press.

Napier, A.S. 1906. "Contributions to Old English Lexicography". *Transactions of the Philological Society* 25 (3). 265-335.

Phillipps, Thomas. 1838. *Fragment of Ælfric's Grammar, Ælfric's Glossary, and a poem on the soul and body, in the orthography of the 12th century: discovered among the archives of Worcester Cathedral by Sir T. Phillipps*. London: W. Clowes and Sons.

Pulsiano, Phillip. 2000. "William L'Isle and Editing". In: *The Recovery of Old English: Anglo-Saxon Studies in the Sixteenth and Seventeenth Centuries*, edited by Timothy Graham, 173-206. Kalamazoo: Medieval Institute Publications.

Roberts, Jane & Christian Kay. 2017. "The Source Dictionaries". In *About A Thesaurus of Old English.* Glasgow: University of Glasgow. Accessed July 24, 2020. http://oldenglishthesaurus.arts.gla.ac.uk/source-dictionaries.

Rodríguez-Álvarez, Alicia. 2009. "'With a concise historical account of the language': Outlines of the history of English in eighteenth-century dictionaries". In C*urrent issues in late modern English*, edited by Ingrid Tieken-Boon van Ostade and Wim van der Wurff, 183-208. Bern/Berlin: Peter Lang.

Somner, William. 1659. *Dictionarium Saxonico-Latino-Anglicum*. Oxford: excudebat Guliel. Hall, pro authore.

Spelman, Henry. 1626. *Henrici Spelmanni Equit. Anglo-Brit. Archæologus.* London: apud Iohannem Beale.

Stanley, Eric G. 1985. "The Treatment of Late, Badly-Transmitted and Spurious Old English in a Dictionary of that Language". In *Problems of Old English Lexicography: Studies in Memory of Angus Cameron,* edited by Alfred Bammesberger, 331-68. Regensberg: Verlag Friedrich Pustet.

Sweet, Henry. 1892. *A New English Grammar, Logical and Historical, Part I.* Oxford: Clarendon Press.

Sweet, Henry. 1896. *The Student's Dictionary of Anglo-Saxon.* Oxford: Clarendon Press.

Thorpe, Benjamin. 1834. *Analecta Anglo-Saxonica: A selection, in prose and verse, from Anglo-Saxon authors of various ages; with a glossary.*

Designed chiefly as a first book for students. London: John and Arthur Arch.
Toller, T. Northcote. 1921. *An Anglo-Saxon Dictionary: Supplement*. Oxford: Clarendon Press.
Treharne, Elaine. 1997. *The Old English Life of St. Nicholas with the Old English Life of St. Giles.* Leeds Texts and Monographs, New Series 15. Leeds: Leeds Studies in English, University of Leeds.
Treharne, Elaine. 2012. *Living Through Conquest: the politics of early English, 1020-1220.* Oxford: OUP.

CHAPTER SIX

BEYOND THE "GRIMM": GERMAN HISTORICAL LEXICOGRAPHY AFTER *DEUTSCHES WÖRTERBUCH*

VOLKER HARM

Abstract

In 2018, when the last fascicle of the second edition of the *Deutsches Wörterbuch* (DWB) appeared, a great tradition of dictionary making that started with the brothers Grimm in 1838 came to a preliminary end. With 32 volumes of its first edition (^{1}DWB), completed in 1960, and 9 volumes of its second edition (^{2}DWB), the dictionary represents the most encompassing description of the German language. For the historical lexicography of German, the completion of ^{2}DWB marks a turning point and gives rise to a number of questions: Is there still a need for a diachronic description of German vocabulary along the lines of the Grimmian tradition? And if so, what should such a dictionary look like?

In order to answer these questions, this article, first, gives a brief account of ^{2}DWB (which is still less known than it deserves) and, second, introduces the new project *Wortgeschichte digital* as its successor. The fundamentals of *Wortgeschichte digital* are outlined and it is shown in which respect the project can be located within the tradition of ^{1}DWB and ^{2}DWB. Special emphasis is laid on innovative ways of historical lexicography which the project tries to implement, notably on the narrative format of its entries.

1. Introduction

For more than 180 years, German speaking philologists have been working on the *Deutsches Wörterbuch* or the "Grimm", i.e. the dictionary project which was founded by Jacob Grimm und Wilhelm Grimm in 1838 and the first edition of which was completed more than 120 years later, in 1960 (hereafter ^{1}DWB). In 2018, when the last installment of the second edition of *Deutsches Wörterbuch*, the *Neubearbeitung des Deutschen Wörterbuchs* appeared (which will here be referred to as ^{2}DWB), this long-standing tradition came to an end—and we still do not know whether the completion of ^{2}DWB really is the end for German lexicography in the Grimmian style or whether we have only reached a milestone in a still ongoing journey. Be that as it may, the completion of the ^{2}DWB (and its subsequent digitization[1]) undoubtedly marks a turning-point in the history of German lexicography. It is now in a situation in which it has not been for a very long time: It is without an overarching national dictionary project and without a common reference point within the lexicographical landscape (on the present state of German historical lexicography, see also the account of Kirkness 2016).

This new situation gives rise to a couple of questions. The most fundamental question is: Do we still need a historical dictionary in the tradition of DWB, or is the work on the history of German vocabulary after 180 years of lexicographical labor already done? And if not so, how should it best be done: Should it take the form of a third edition of the dictionary—a ^{3}DWB much in the style of the ^{3}OED, the most recent online version of the *Oxford English Dictionary*? Or are completely new ways of doing historical lexicography imaginable?

In order to better understand the current state of German lexicography arising from the completion of ^{2}DWB, we have to step aside for a while and take a closer look at the history and the achievements of Grimmian lexicography. This retrospective will be the first part of this article and it will be concerned with the less-known ^{2}DWB rather than with ^{1}DWB. In the second part of the article the question of how to proceed after the end of the ^{2}DWB will be addressed. And in the third part, a new project in the Grimmian tradition will be presented: *Wortgeschichte digital*.

2. The lexicography of ^{2}DWB: a retrospective

A brief sketch of ^{2}DWB and its history is not only necessary in order to understand the present state of dictionary-making in the German speaking countries; but it is also useful for its own sake because ^{2}DWB is still less popular than ^{1}DWB. This first edition, as indicated above, was founded by Jacob Grimm and Wilhelm Grimm in 1838 and was completed in 1960 after

[1] Since 2020, the digitization of ^{2}DWB is available on: https://www.dwds.de/d/wb-2dwb and https://woerterbuchnetz.de/?sigle=DWB#1, ^{1}DWB is on: https://woerterbuchnetz.de/?sigle=DWB.

a difficult and complicated history[2]. While working on the last volumes of ^{1}DWB in the late 1950s plans were already made to continue the project. Two alternatives were under discussion: The first option—informally termed "Kurz-Grimm"—was a short edition of ^{1}DWB, reducing the original 32 volume edition into an updated version of more or less 10 volumes. The second option was a revision focusing only on the oldest parts of ^{1}DWB ("Reparaturlösung"), leaving the more recent volumes untouched, at least for the time being. After an intense debate the second solution won out: The Berlin and Göttingen Academies of Sciences which had responsibility for the project decided to prepare an entirely new edition of the most outdated volumes of ^{1}DWB; the project was referred to as *Neubearbeitung des Deutschen Wörterbuchs von Jacob Grimm und Wilhelm Grimm* (short: ^{2}DWB). Ironically, the most outdated volumes of ^{1}DWB were those which had been compiled by the Grimms themselves, i.e. the volumes 1 to 4 ranging from the letter A to F. In consequence, the ^{2}DWB, as it were, is nothing else than the expulsion of the founding fathers from their own dictionary. But, after all, the idea of a new edition of this part of the dictionary made sense: The volumes compiled by the Grimms date from the early days of German philology; they are relics of the heroic age of our discipline, but cannot be consulted successfully whenever reliable and up-to-date information on word usages or word histories is needed. Jacob Grimm himself, when compiling the dictionary, was already aware of the fact that the completed volumes needed revision: "bücher dieser art können erst gut werden bei zweiter auflage" ("books of this kind can only become real good ones with their second edition", Jacob Grimm to Karl Weigand, see *Briefwechsel* 382).

The dictionary scheme for the second edition which was made in the late 1950s included the following key points: 1) The four ^{1}DWB volumes A-F should be replaced by nine to ten volumes of ^{2}DWB. 2) The new volumes should be worked out in the same way as the most recent volumes of ^{1}DWB, i.e. the volumes of the period from 1930 to 1960. This should give the dictionary a more homogeneous shape. 3) Rather than a revision ("Neubearbeitung") in the proper sense of the word, the ^{2}DWB was designed as an entirely new dictionary based on a new collection of quotation slips. 4) The dictionary's coverage should be more strictly confined to the period from approximately 1450 up to the present; earlier language stages, i.e. Old and Middle High German, are only included insofar as they illuminate the history of words younger than 1450. Limiting the dictionary's scope implied also a return to the Grimm brothers' original idea of a

[2] For more details on the history of ^{1}DWB, see Kirkness (2012) and Harm (2020).

dictionary ranging “from Luther to Goethe” (the latter here standing for the present-day language of the Grimm’s lifetime). 5) The entries are focused on historical semantics rather than on encyclopedic or stylistic features of word usage, which were abundantly described in some of the earlier ^{1}DWB volumes.[3]

The similarities and differences between ^{1}DWB and ^{2}DWB become evident when comparing any page of both editions. Here, the entry *Gabel* from ^{1}DWB, dating from 1872, and *Expedition* from ^{2}DWB, published in 1999, have been selected (see Figs. 6-1 and 6-2, respectively).

3) *von allerhand ähnlichen dingen.*

a) gewachsene gabel *von holz u. ä.* *α) gabelförmiger ast oder zweig, auch* zwiesel (*von den zwei armen*), *s. z. b.* eichengabel, *noch am baume. auch botanisch, cirrus. ähnlich* gablen *pl. an weinreben, s.* MAALER, RÄDLEIN *und* gäbelein 2.

β) dann bes. ein solches gabelstück im gebrauch des lebens, als stütze (*lat.* furca, cervus), *z. b. für die musketen alter zeit, für schwere fahnen* (*s.* gabelmann), *für jagdnetze* (jegergabel DIEF. *s. v. amis,* netzgabel HENISCH), *für geländer* (*auch* furkel), *in booten als unterlage und halt der ruder; auch zum heben* (*vergl.* hebegabel); *weidm. zum niederdrücken eines gefangnen wilden thiers am halse; beim vogelfange, ein birkenreis, finken aufs stechen zu fangen, mhd.* vogelgabel *hamus. s. auch* furkel.

γ) dasz diesz nicht die ursprüngliche bed. ist, sondern erst nach der unter 1 *entstanden, ist an sich notwendig, da man die dinge nur nach ihrem gebrauch nannte und zu nennen anlasz hatte. daher auch das fremdwort* furkel, *das mit fremdem jagdgebrauch u. ä. ins land kam.*

b) die deichsel eines einspänners. auch sonst ähnliche einfassende werkzeuge und werkstücke in mancherlei gewerken. bei HENISCH *kelterscheer, furca s. cornua praeli.*

c) das hirschgeweih, wenn die stange nur ein ende trägt, vgl. gabeler; *auch die beiden obersten enden des geweihs. vgl.* forkel 1 *und* forkeln.

d) gabelförmiger knochen, so beim geflügel an der brust. ähnlich wol am menschen: die lacerti (*muskeln*), die das haupt bewegen, seind zweier hand. etlich bewegen daz haupt on die anderen und haben iren ursprung hinder den oren, bitz sie kummen zů der gabelen. GERSDORF 3.

Fig. 6-1: Entry *Gabel* ‘fork’ from ^{1}DWB 4, 1121 (extract).

[3] On basic features of ^{2}DWB see also the preface in vol. 1, 3-8.

EXPEDITION *f. lehnwort aus lat.* expeditio *f.* 3 *wohl im anschluß an engl.* expedition *subst.* **1** *das fortziehen in andere, meist fremde gebiete oder orte; metonymisch auch die gruppe der fortziehenden.* **a** *kriegszug, eroberungszug:* 1509 wo auch wider die unglaubigen oder Turcken mit einer stattlichen expedicion oder zugk sollt gehandelt werden .. so .. zuvor .. geratschlag und ermessen wurde, wie und welcher massen solicher zugk .. zum besten und geschicksten sollt .. werden *Frankf. reichscorr. 2,2,770 J.* 1633 nach dieser .. sieghafften expedition ist der kayser sampt dem heer .. wider in Hispanien geschiffet GOTTFRIED *hist. chr. (1630)7,58.* 1767 Essex kömmt von seiner expedition wider die Spanier zurück, und will der königinn in London bericht davon abstatten LESSING *10,34 L./M.* 1838 die expedition Mehmet=Paschas besteht aus drei bataillonen des ersten und drei des zweiten linien=infanterie=regiments MOLTKE *ges. schr. (1891)8,270.* 1964 Großbritannien solle künftig .. militärische expeditionen und die bombardierung von ortschaften, unterlassen *Keesing's arch. 34,11071^{b}.* **b** *unternehmung, fahrt, ausflug, seit dem 18. jh. besonders forschungsreise in unentdeckte gebiete:* 1566 wie aber der doctor .., als er wider haim kommen, seiner expedition halb beim herzogen bestanden, das hab ich nie .. erfaren *zimmer. chr. ^{2}2,217 B.* ⟨1618⟩ nach erlangung hoher aempter haben sie *(landes und allgemeine friedens feinde)* .. bey königlichen vnd landts expeditionen .. alte vnnd nützliche diener schimpfflich abgeschafft *in:* ENSZ *fama (1627)87^{a}.* 1785 die meisten dörfer .. liegen nahe um Göttingen .. und .. machen die gränzen für die aus, welche dergleichen expeditionen zu fuß unternehmen LIST *statistik 16.* 1899 der

Fig. 6-2: Entry *Expedition* from ^{2}DWB 8, 2507 (extract).

What is striking in the first place is that ^{2}DWB, as illustrated by the entry *Expedition* in Fig. 6-2, keeps using the orthography introduced by Jacob Grimm in which, in contrast to German spelling customs from the sixteenth century up to the present, the nouns were not capitalized. Around 1960, the Grimmian spelling is a significant anachronism since Jacob Grimm never succeeded in imposing his personal spelling rules on German orthography. Apparently, the decision to leave the orthography of ^{1}DWB unchanged was taken mostly for tactical reasons: Since it was difficult to explain to a wider audience why and in which respect ^{2}DWB is still "the Grimm", this visual feature made it clear that the dictionary really stands in the venerable line of tradition reaching back to the founders. That the brothers are still a prominent part of the title of ^{2}DWB points in the same direction.

There were, however, several innovations and deviations from the ^{1}DWB tradition hiding behind the dictionary's nineteenth century garment.

One entirely new feature catching the eye here is the dating of the quotations (see Fig. 6-2), which was not common even in the most recent volumes of ^{1}DWB. The two example entries *Gabel* and *Expedition* also reveal a second and very significant difference between ^{1}DWB and ^{2}DWB: The ratio of descriptive text and quotations has changed insofar as definitions and explanations are much shorter in the second edition than in the first; in ^{2}DWB, much more space is given to the quotation evidence. This goes hand in hand with a flatter hierarchy of the entry structure (*Expedition* in Fig. 6-2 has two hierarchy levels, as opposed to three in *Gabel* and five levels, e.g., in the entry *stehen* [^{1}DWB 17, 1473]). Accordingly, the entries of ^{2}DWB are no longer labyrinthine monographs with an intricate structure, they rather consist mostly of only two or three hierarchy levels. The features of ^{2}DWB mentioned here—dating of the quotations, stronger focus on the evidence, a clear entry structure—are very much reminiscent of the OED, and it is highly probable that at this point of the history of DWB, the OED indeed has exerted some influence.[4]

When the ^{2}DWB was planned, it was estimated that it would take approximately twenty years to complete the 10 volumes ranging from A to F. This was yet another gross miscalculation in the history of the DWB, which is abounding in miscalculations and overestimations from its very beginnings. The compilation of ^{2}DWB took nearly three times longer than originally planned: 58 years passed between the publication of the last volume of ^{1}DWB (in 1960) and the publication of the last installment *Büffelkuh–Cyberspace* in 2018. There are mainly two reasons for this slow progress: The first reason is undoubtedly the high quality of the dictionary. It was compiled on the basis of approximately 5.5 million quotation slips for the first six letters of the alphabet. Compared to the ^{1}DWB, which disposed of approximately 2.5 million quotations for the whole dictionary from A to Z, this is a significant enhancement of the dictionary's empirical basis (for the figures, see Harm 2020, 73-75). The second reason for the slow progress of ^{2}DWB is of a political nature. As an exceptional project of both German states during the period of the Iron Curtain, it faced substantial difficulties under the communist regime. In East Berlin, many of the ^{2}DWB lexicographers were transferred to other projects which were thought to be more useful for socialism, particularly to the *Wörterbuch der deutschen Gegenwartssprache* (WDG).[5] As a consequence, when the German Democratic Republic collapsed in 1990, the part taken over by the East

[4] This has to be shown in more detail in a special research project on the history of ^{2}DWB which the author is currently conducting.

[5] The WDG, which appeared from 1961 to 1971 (6 volumes, 1st edition), is the main basis for the *Digitales Wörterbuch der deutschen Sprache* (DWDS), www.dwds.de.

German *Akademie der Wissenschaften der DDR* had not reached the end of the letter A.

3. German lexicography after the completion of ^{2}DWB: an aporia

When the work on ^{2}DWB began it was considered only as the first and most urgent step in an encompassing revision of the whole dictionary. Today, knowing how difficult and time-consuming the work on the new edition of A-F really was, there is no question of such a revision for the letter section G-Z anymore, at least not in the manner of ^{2}DWB. Without a dramatic increase in resources (which is not in sight), it would take decades, if not a century to proceed in the way ^{2}DWB has been compiled. Therefore, a continuation of the work in the style of the ^{2}DWB is out of the question.

But before considering the issue of what to do with the Grimm tradition in the future, the more fundamental question has to be answered as to whether we still need such a dictionary at all. In this case, it is not difficult to give a positive answer: There are many good reasons for such a dictionary. I will just mention one: The DWB is the only dictionary treating the more recent history of German, i.e. the period after 1600 up to the present; this period is covered by no other historical dictionary in a sufficient way. And this period is important as it forms the missing link between our linguistic past after the era of Renaissance and Reformation on the one hand and our present-day language on the other. For this reason, it is evident that the DWB has to be continued in one or another way—unless we do not want to content ourselves with a state of knowledge dating from the late nineteenth or mid-twentieth century.

The more difficult question is how to do so: How could the DWB possibly be continued? The first idea that comes to mind is a revision in the style of the third edition of the *Oxford English Dictionary*. In such a revision of ^{1}DWB—which would be an online dictionary, of course—the most outdated volumes would be incrementally edited by reworking obsolete passages and by adding new quotation evidence, new meanings and new entries. One important task of such a revision would be to provide all quotations with accurate dates which ^{1}DWB lacks in contrast to ^{2}DWB, see Figs. 6-1 and 6-2. (In the course of such a revision, by the way, one could also take the opportunity to delete the few Nazi quotations which disfigure the volumes published between 1933 and 1944—this is a wish which has been very frequently uttered by users of ^{1}DWB; for more information, see Schmitt 2010).

The example of [3]OED, which undoubtedly is a great lexicographical achievement and which has to be considered as a model unlike any other, suggests that this kind of revision could be a feasible way for DWB, too. When looking at the details, however, it becomes evident very quickly that such a revision would cause a couple of serious problems. One obvious difficulty with [1]DWB is that, due to its long-standing history, it can hardly be considered as one dictionary: Many volumes, notably those from the second half of the nineteenth century, were compiled by lexicographers who worked almost independently and without coordination and supervision. [1]DWB, therefore, is far more heterogeneous than the OED, and, accordingly, revising it would be quite a laborious task. A second characteristic of [1]DWB which makes a wholesale revision almost impossible is that the implementation of the so-called historical principle of lexicography (see Considine 2016, 168-169), i.e. the chronological arrangement of subsenses and (dated) quotations, which is standard in OED and [2]DWB, is only one structuring principle among many others in [1]DWB. Applying the historical principle more or less consistently to the whole dictionary would amount to a far-reaching reorganization of many entries of [1]DWB—not a few entries would have to be turned upside down. The implementation of the chronological arrangement of subsenses and quotations requires another round of fundamental work: The roughly two million quotations of [1]DWB have to be provided with a date, i.e. with an indication of the year or the timespan in which the text was composed or published. This task is far from trivial, as a great part of the quotations in [1]DWB, particularly if taken from editions, has to be looked up in the original text—provided that the quotation can be identified at all, which is a problem for many bibliographical references from the nineteenth-century volumes. Therefore, dating the quotations would be an extraordinarily time-consuming task. And even if it could be done, it is questionable whether the result would be worth the effort: Entries from the nineteenth century with dated quotations and a chronological rearrangement of senses still remain entries from the nineteenth century. Even if dated quotations were eventually available, the enormous work of revising and modernizing the entries of [1]DWB still remains to be done. In consequence, proceeding this way, only a few entries could properly be revised. For this reason, it seems more reasonable to compile an entirely new dictionary instead of revising [1]DWB with huge efforts and without certainty as to the quality of the results.

4. A new project in the Grimmian tradition: *Wortgeschichte digital*

4.1 A dictionary project in its context

At present, historical lexicography of German is in a paradoxical situation. As just demonstrated, a revision of ¹DWB along the lines of the ³OED is out of the question; a new dictionary would be better. ²DWB was an entirely new dictionary, replacing the letter section A-F of ¹DWB, but its completion took far too long, so ²DWB is not a model either. How to overcome this apparent aporia—how could we develop a new dictionary in the Grimmian tradition while at the same time avoiding the pitfalls of ²DWB?

This article would not have been written if there were no solution within reach. At the beginning of the year 2019, a new dictionary project started at the *Göttingen Academy of Sciences and the Humanities*, one of institutions responsible for ²DWB. The project is called *Wortgeschichte digital*. It is part of the *Zentrum für digitale Lexikographie der deutschen Sprache* (ZDL), a new research center which has been founded by the Academies of Berlin and Göttingen, together with the *Institut für deutsche Sprache* in Mannheim and other German Academies involved in historical lexicography (*Akademie der Wissenschaften und Literatur Mainz* and *Sächsische Akademie der Wissenschaften*). The new center aims at coordinating the digital lexicography of the German language, including corpus research and IT-infrastructure, and it is concerned with present-day language as well as with language history.

Wortgeschichte digital (WGd), the project which will be presented in the following section, is the historical part of the new lexicographical center.[6] WGd has the task of compiling a period dictionary covering the most recent stage in the history of the German language, i.e. New High German (from 1600 to the present). For this period, no other dictionary exists so far: Early New High German (from 1350 to 1600/1650) is covered by *Frühneuhochdeutsches Wörterbuch* (FWB), and for the present-day language, we have *Duden* and *Digitales Wörterbuch der deutschen Sprache* (www.duden.de and www.dwds.de). WGd, therefore, fills a significant gap within the historical lexicography of German.[7]

[6] For the time being, the duration of the project, as of the whole lexicographical center, is five years. During this period, the technical and lexicographical basis of the project will be created; after this initial phase, WGd and ZDL will be established as permanent institutions. Currently, five lexicographers are working on the project. For further information, see http://wdg.adw-goe.de/.

[7] For more details, see Harm (2019).

But how exactly does WGd fit in the line of tradition of *Deutsches Wörterbuch* and notably of ²DWB? As indicated previously, WGd covers the most recent stage of German language history, i.e. New High German, up to the present day. This connection of the linguistic present with the more recent past has to be considered as a hallmark of Grimmian lexicography. Rather than compiling a historical dictionary in the proper sense of the word, Jacob and Wilhelm Grimm intended to give an encompassing description of the modern German language of the last four hundred centuries. Their dictionary intends to connect and even to unify the linguistic past and present of German, looking for ways to make the linguistic past interesting for the audience of their time.[8]

The second core feature WGd shares with DWB is its diachronic approach. This means that WGd will not treat the period from 1600 to the present as a synchronic unity in the way of a prototypical period dictionary like *Mittelhochdeutsches Wörterbuch* (MWB) or *Frühneuhochdeutsches Wörterbuch* (FWB). WGd, thus, does not confine itself to listing senses according to general criteria like general vs. specific or concrete vs. abstract. Rather, it will focus on the evolution of word meanings towards the present, trying to reconstruct the evolving network of word senses during the last four hundred centuries. This diachronic, i.e. truly historical, approach is one reason why the project is called "Wortgeschichte" ("word history").[9]

A third aspect in which WGd continues the tradition of the DWB concerns the historical embedding of word histories. ¹DWB and notably ²DWB are concerned with the language from 1450 onward, but normally include a brief sketch of earlier stages of a word's history (if extant). In this respect, too, ¹DWB and ²DWB differ significantly from period dictionaries which normally do not look backwards into previous language history. For WGd, the presentation of a word's earlier history much in the way of ²DWB is an important feature. By providing an overview of those earlier word senses which are relevant for the understanding of the more recent sense developments, WGd makes its word histories more comprehensible.

[8] For this characteristic of ¹DWB, see Jacob Grimm's preface in ¹DWB vol. 1, p. xii, xix: According to J. Grimm, the dictionary is a "denkmal des volks, dessen vergangenheit und gegenwart in ihm sich verknüpfen" (¹DWB 1, xii). "a sublime monument of the people, whose past and present are connected with each other".

[9] The question as to whether ¹DWB really is "diachronic" in the strict sense of de Saussure, is difficult to answer as it implies a serious anachronism. There is no general answer for the whole dictionary; on the question of a "diachronic" approach in the volumes of the brothers Grimm, see Harm (2014).

4.2 Innovations in WGd against the background of DWB

As these features—connection of linguistic past and present, diachronic approach, inclusion of earlier stages—have been core characteristics of DWB since the days of the brothers Grimm, WGd can be considered as the legitimate successor of the Grimmian tradition. But in other respects, it develops and modifies this tradition in different directions. The way in which it differs from DWB most significantly concerns its basic approach to vocabulary. Whereas DWB worked alphabetically, proceeding from word to meaning in the classical semasiological way of dictionary making, WGd is onomasiologically based. Its lemma selection and linguistic description starts not from words, but from conceptual domains such as, e.g., society, politics, economy, technology, communication, traffic, everyday culture, thereby preferring those domains of vocabulary which underwent significant change from 1600 to the present. The onomasiological approach chosen by WGd has many advantages over the classical method: Instead of arbitrarily shifting from topic to topic following the contingencies of the alphabet, the lexicographers can apply specialized focus on selected domains in a very efficient way; the words belonging to a common field can be described in connection to each other, so that the user gets a far more coherent picture of the lexical developments.

The second way in which WGd deviates from ²DWB is the style of presentation. This has to be explained in some detail. As is well known, the traditional way of describing a word history consists of a combination of definitions and quotations, both being arranged in chronological order. This manner of presentation—often termed "the historical principle", as mentioned above—goes back to the German lexicographer Franz Passow, who stated in the preface to his *Handwörterbuch der griechischen Sprache* (1824), "that every word should be made to tell its own story" (quoted from Zgusta 2006, 29). The Passowian assumption that every word is able to tell its own story has a strong positivistic propensity, suggesting that the sources themselves give their own history. But of course, words do not tell stories on their own, and a mere arrangement of quotations and the chronologically ordered listing of word meanings is not yet a word history. It is the lexicographer who has to do that work by interpreting historical material and by construing semantic filiations and developments. In WGd, we therefore aim at making these constructions more explicit than in previous historical dictionaries. Therefore, a word's history will be told in a short text rather than being hidden somewhere in the structure of the entry. This kind of presentation is not very common in lexicography (except, of course, in encyclopedias), but there are a few models we can lean on, notably Alain Rey's *Dictionnaire de la language française* (DHLF) and Raymond

Williams' *Keywords* (Williams 1976). Since [1]DWB, especially in its older volumes, also shows a tendency towards a more narrative style of lexicography (which can be illustrated by the entry *Gabel* in Fig. 6-1, too), the word histories of WGd can be considered in some way as a return to a Grimmian tradition that has been lost in [2]DWB. By choosing a narrative style, WGd even takes up a core idea of Jacob Grimm, who conceived [1]DWB as a "Hausbuch", i.e. as a book that can be read and understood not only by scholars, but also by average educated people ([1]DWB 1, xiii). To make the dictionary accessible to a larger audience is an important motivation for the decision to write a readable text instead of preparing a complex and in many cases rather arcane microstructure.

What such a word history looks like might well be illustrated by an extract of the WGd entry *Team* (a loanword from Engl. *team*).

> *Erste Entlehnung aus dem Englischen*
>
> Das Wort stellt eine Übernahme von englisch *team* dar (zum etymologischen Zusammenhang mit dt. *Zaum* s. Pfeifer). Dabei werden unterschiedliche Bedeutungen des englischen Wortes zu unterschiedlichen Zeitpunkten ins Deutsche entlehnt: Zunächst findet engl. *team* in der Bedeutung ‚Gespann von Zugtieren' Eingang in deutsche Texte (1902). In dieser Lesart bleibt es im Deutschen aber insgesamt selten (so etwa noch 1936).
>
> Team *als Ausdruck des Sports*
>
> Wichtiger für die Geschichte des Wortes im Deutschen ist die Verwendung von *Team* als Ausdruck des Sports: Engl. *team* bedeutet ‚Mannschaft, die an einem sportlichen Wettkampf teilnimmt'. Das erste Zeugnis für die Kenntnis dieses englischen Wortes im Deutschen ist eine von Konrad Koch, dem Begründer des Fußballspiels in Deutschland, aufgestellte Liste mit Verdeutschungsvorschlägen für englische Ausdrücke des Fußballsports. Als Ausdrücke, die engl. *team* ersetzen sollten, werden die Wörter *Mannschaft* und *Riege* vorgeschlagen (1903). […]
>
> Trotz der sprachpflegerischen Bemühungen Kochs um einen Ersatz von *Team* durch *Mannschaft* und *Riege* findet sich das Wort in der Folgezeit häufig im Bereich des Sports, und zwar nicht nur mit Bezug auf englisch-amerikanische Verhältnisse, wie im „Deutschen Fremdwörterbuch" angedeutet wird, sondern durchaus auch für heimische Sportmannschaften (1908, 1911).[10]

[10] Translation:

Although WGd presents word histories in a narrative format, these histories are worked out in the traditional philological way, i.e. by interpreting and arranging a large amount of quotation evidence which in the case of our project is largely taken from *Deutsches Textarchiv* and the corpora of *Digitales Wörterbuch der deutschen Sprachen* (DWDS).[11] A relevant part of these quotations has to be made accessible to the dictionary's users in order to enable them to verify and criticize the description. For this purpose, a quotation section is presented alongside the text. What is most important is that the relevant historical statements within the text are linked directly with the quotations which form their empirical basis (see Fig. 6-3). Hereby, the user can read the text without being interrupted by the quotations, but, at the same time, s/he can verify the statements made in the text. So, the quotation evidence which is the bedrock of scientific lexicography ever since the nineteenth century is not given up; rather, its function and its connection with the lexicographical description is made much more explicit as the reader is not compelled to search the whole quotation section for the relevant example.

Early loans from English

The word is a borrowing from English *team* (for the etymological connection with German *Zaum* see Pfeifer). Different senses of the English word are transferred into German at different times. First, *team* in the meaning 'two or more horses or other animals harnessed together to pull a vehicle' is used in German texts (1902). In this reading, the word is only scarcely attested (see e.g. 1936).

Team *as an expression of sports*

More important for the history of the word is the domain of sports: Here, English *team* means 'a number of players forming one side in a competitive game or sport'. The first attestation of this usage in German is in a list of translations for English football terms which was compiled by the football pioneer Konrad Koch. As indigenous alternatives for the English expression, Koch proposes the German words *Mannschaft* and *Riege* (1903). […].

In spite of Koch's attempt to replace the English word by *Mannschaft* or *Riege*, *team* was more and more established in German texts dealing with sports, not only with respect to Anglo-American events, as *Deutsches Fremdwörterbuch* holds, but also with respect to domestic referents. […]

[11] For the project, Nico Dorn has developed the app *Zettel's Traum* which allows lexicographers to organize a large amount of quotations and to easily prepare the online presentation of the entries. The app includes, e.g., an import function for quotations from a corpus, export to XML, lexicographical ascriptions, filters etc. The app is available on https://github.com/WortgeschichteDigital/ZettelsTraum.

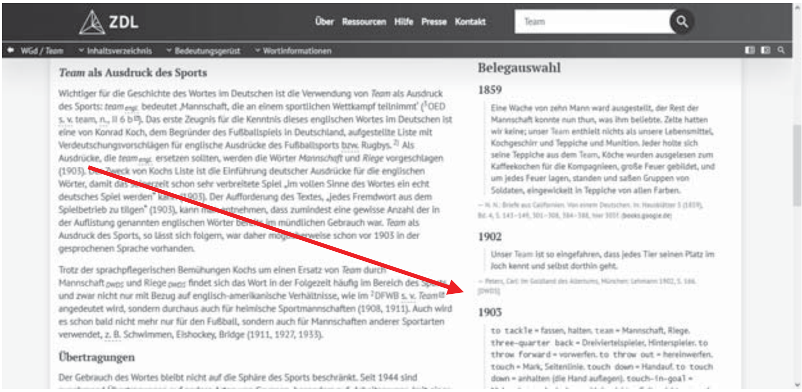


Fig. 6-3: Word history *Team* (WGd) with internal link to the quotation section.

This narrative structure is mainly designed for users who have a basic interest in word history and who are willing to take the time to read such an entry or at least relevant parts of it (many of the labyrinthine entries of [1]DWB and [2]DWB, by the way, appear to address a very patient reader, too, cf. [1]DWB 1, xiii). There are other kinds of users, of course, who just want to have a brief answer to a very specific question. And there are, thirdly, users who want to search the dictionary as a database looking for answers on questions beyond a single word (e.g. "Give me all the loanwords from English within the domain of sports which entered German in the second half of the nineteenth century"). As the dictionary has to meet all these expectations, the word histories—which are nonetheless the core of the project—have to be accompanied by additional representations. For this purpose, a short abstract of the word history (see "abstract" / "Kurz gefasst" in Fig. 6-4) and an overview of those word meanings which play a role in the word history are provided (see "Bedeutungsgerüst" / "relevant senses" in Fig. 6-4). This overview allows the reader to look up, e.g., just a meaning s/he does not understand. The list of senses is linked with the relevant passage in the text which contains more information on historical aspects of the usage in question; so, the user who starts with the overview of senses has direct access to the piece of information s/he might be interested in. In order to impose a structure on the text which, in many cases, consists of more than one page, the word history is divided into small chapters or sections much in the way of Wikipedia entries; the titles of these sections, which are presented at the beginning, are linked with the content (see also Fig. 6-4). In order to make the text searchable, it will be tagged at least to a certain extent; so that the user can search the whole dictionary, e.g., for "loanword" or "metaphor". But, as WGd started only in 2019, many technical details have still to be determined.

It goes without saying that all the features presented here are only possible in a digital format (therefore "Wortgeschichte digital"). WGd, thus, takes full advantage of the computer as a working tool and of the internet as a medium of publication.[12] Therefore, no print edition is planned, at least not for the moment. In this respect, WGd breaks with Grimmian tradition perhaps in the most obvious way.

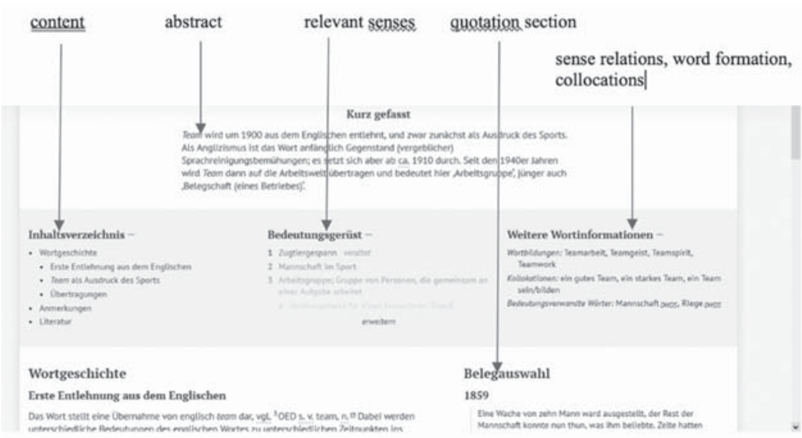


Fig. 6-4: Additional components of a WGd entry.

5. Conclusion

At present, historical lexicography of German finds itself in a period of transition: With the completion of the second edition of *Deutsches Wörterbuch*, which only extends from A to F, a line of tradition dating from the first half of the nineteenth century comes to an end and the question arises how to proceed with lexicography in the Grimmian tradition. A revision of other [1]DWB volumes is not an option, as [1]DWB is too heterogeneous and, in many of its volumes, too far away from the standards of historical lexicography of the twenty-first century. Rather than investing a large amount of work into a revision which, if ever finished, would hardly be satisfactory, it has been decided to chart a new course and to compile a

[12] For more details, see the project site on http://wdg.adw-goe.de/ and https://www.zdl.org/wb/wortgeschichten.

new dictionary of the most recent language period—New High German (from 1600 to the present)—thereby filling the most urgent gap in the historical lexicography of German. This new project, *Wortgeschichte digital*, in spite of its innovations, adheres to DWB as to some of its central hallmarks: It adopts a diachronic approach, it connects linguistic past and present-day language, it presents word histories within their context by sketching their history before 1600. Even WGd's most significant innovation, i.e. the narrative format of the entries, links up with certain traditions of dictionary making in [1]DWB, notably with Jacob Grimm's idea of a dictionary for a broader audience. That we are breaking new ground by making the "Grimm" an online dictionary is not in contradiction with this tradition: The Grimms themselves have always been open to new ways of doing lexicography, and if there is a characteristic of their work, it is a constant and sometimes radical renewal and revision of their own methods and results. Therefore, we are quite confident, that they would have approved of our way of continuing the work they began more than 180 years ago.

Bibliography

Dictionaries

DHLF = Rey, Alain. *Dictionnaire historique de la langue française*. 1992, vols. 1-2. Paris: Le Robert.

[1]DWB = Grimm, Jacob, and Grimm, Wilhelm. 1854-1971. *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm*, vols. 1-16 [1-32]. Leipzig: Hirzel (https://www.woerterbuchnetz.de/DWB1)

[2]DWB = Grimm, Jacob, and Grimm, Wilhelm. 1960-2018. *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm. Neubearbeitung*, vols. 1-9. Hrsg. von der Berlin-Brandenburgischen Akademie der Wissenschaften (vormals Akademie der Wissenschaften der DDR) und der Akademie der Wissenschaften zu Göttingen. Leipzig/Stuttgart: Hirzel. (https://www.dwds.de/d/wb-2dwb, https://www.woerterbuchnetz.de/DWB2).

MWB = *Mittelhochdeutsches Wörterbuch*. Im Auftrag der Akademie der Wissenschaften und der Literatur Mainz und der Akademie der Wissenschaften zu Göttingen herausgegeben von Kurt Gärtner, Klaus Grubmüller, Jens Haustein und Karl Stackmann. Stuttgart: Hirzel 2006, vol. 1-.

FWB = *Frühneuhochdeutsches Wörterbuch*, herausgegeben im Auftrag der Akademie der Wissenschaften zu Göttingen von Ulrich Goebel, Anja

Lobenstein-Reichmann, Oskar Reichmann, begründet von Robert R. Anderson, Ulrich Goebel, Oskar Reichmann. Berlin and New York: de Gruyter 1986, vol. 1-.

Williams, Raymond. 1976. *Keywords. A vocabulary of culture and society*. New York: Oxford University Press.

WDG = Klappenbach, Ruth and Wolfgang Steinitz. *Wörterbuch der deutschen Gegenwartssprache*. 1964-1976, Berlin: Akademie, vols. 1-6.

Secondary sources

Briefwechsel der Brüder Jacob und Wilhelm Grimm mit Rudolf Hildebrand, Matthias Lexer und Karl Weigand, edited by Alan Kirkness. Stuttgart: Hirzel 2010.

Considine, John. 2016. "Historical Dictionaries. History and Development; Current Issues". In *The Oxford Handbook of Lexicography*, edited by Philipp Durkin, 163-175. Oxford: Oxford University Press.

Harm, Volker. 2014. "Wortgeschichte, Etymologie und Organismuskonzeption im Grimm'schen Wörterbuch". In *vnuornemliche alde vocabulen—gute, brauchbare Wörter. Zu den Anfängen der historischen Lexikographie*, edited by Michael Prinz and Hans-Joachim Solms, 257-267. *Zeitschrift für deutsche Philologie* 132. Sonderheft.

Harm, Volker. 2019. "Wortgeschichte digital. Aufgaben und Umsetzungsstrategien". In *Paradigmenwechsel in der Lexikographie – Herausforderung und Chance. Vorträge zum Auftakt des Zentrums für digitale Lexikographie (ZDL)*, herausgegeben von den Präsidenten der Berlin-Brandenburgischen Akademie der Wissenschaften und der Akademie der Wissenschaften zu Göttingen. Berlin: Berlin-Brandenburgische Akademie der Wissenschaften, 47-54.

Harm, Volker. 2020. "Between Science and Romanticism: The *Deutsches Wörterbuch* of the Brothers Grimm". In *The Whole World in a Book. Dictionaries in the Nineteenth Century,* edited by Sarah Ogilvie and Gabriella Safran, 73-92. New York: Oxford University Press.

Kirkness, Alan. 2012. "Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm". In *Große Lexika und Wörterbücher Europas. Europäische Enzyklopädien und Wörterbücher in historischen Porträts*, edited by Ulrike Haß, 211-232. Berlin and Boston: De Gruyter.

Kirkness, Alan. 2016. "Es leben die Riesenschildkröten! Plädoyer für die wissenschaftlich-historische Lexikographie des Deutschen". In *Lexicographica*, 32: 17-137.

Schmitt, Peter. 2010. "Das Wörterbuch im Dienst der Propaganda. Das Deutsche Wörterbuch von Jacob Grimm und Wilhelm Grimm und Trübners Deutsches Wörterbuch". In *Wissenswelten: historische Lexikografie und Europäische Aufklärung*, edited by Robert Charlier, 133-152. Hannover: Wehrhahn.

Zgusta, Ladislav. 2006. *Lexicography Then and Now. Selected Essays*, edited by Fredric S.F. Dolezal and Thomas B.I. Creamer. Tübingen: Niemeyer.

CHAPTER SEVEN

OLD ONOMATOPOEIA: WHAT ETYMOLOGICAL DICTIONARIES TELL US ABOUT SOUND IMITATION IN EXTINCT LANGUAGES

MARIA A. FLAKSMAN

Abstract

Every modern language has a small number of imitative (onomatopoeic, mimetic) words in its lexicon. These words are exceptional as they possess an associative link between form and meaning making the two sides of the linguistic sign almost identical. After Charles Peirce, such property is called ‘iconicity’, and such words are referred to as ‘iconic’. Modern iconic words come up mostly in expressive speech where there is a need for a description of sound phenomena (consider, for example, English *plop*!, *swish*! *hiss*, or Russian *жужжать* ‘to buzz’, *щебетать* ‘to chirp’, *чик-чирик* ‘chirp-chirp’). Together all iconic words of a language make up its (sound) imitative system. The problems addressed in the present research are (1) what information about the older strata of imitative lexicon can be obtained from etymology dictionaries and (2) whether the imitative system of a language remains stable from a diachronic perspective or evolves over time. This research is conducted on the material of the Old English language taken from the Holthausen’s *Altenglishes Etymologishes Wörterbuch*. The article also discusses the problem of application of the methods of phonosemantics (iconicity studies) to the material of ancient and reconstructed languages.

1. Introduction

The growing bulk of evidence from modern languages of different families (see e.g. Hinton et al. 1994; Anderson 1998; Voeltz et al. 2001; Voronin

2006; Dingemanse 2012; *Iconicity Atlas* 2019; Wescott 1980) suggests that onomatopoeic (more broadly—'iconic') words might be a language universal. This gives us a reason to suspect that they might be found in ancient and reconstructed languages as well. Indeed, there are several works devoted to diachronic studies of phono-symbolism (Malkiel 1990; Mithun 1982; Afanasiev 1984; Carling and Johansson 2014), confirming this suggestion. But do these words differ from modern sound imitations? If yes, in what way? For how long do they stay in a language? Do they change? Do they become obsolete?

The purpose of this article is to suggest some possible answers to these questions as well as to outline the general tendencies of expressivity loss in the onomatopoeic lexicon as observed in the material provided by etymology dictionaries (Holthausen 1974; Lehmann 1986; Kroonen 2013; Kluge 2011).

The first part of the chapter is devoted to the discussion of the lexicology-oriented methods of phonosemantics (iconicity studies), while the second part is devoted to the application of these methods and on the research on imitative words in the Holthausen's *Altenglishes Etymologishes Wörterbuch* (1974).

2. Onomatopoeia, sound symbolism, iconicity—setting the terms apart

When it comes to iconicity studies (originally, a sub-branch of semiotics) one has to face the problem of the absence of well-established terminology in the field. All words exhibiting a form-meaning link are generally labelled 'iconic' after Charles Peirce ([1867] 1982), or (also) 'imitative'—as opposed to 'arbitrary' or 'symbolic' (after Saussure 1916). However, one may also encounter terms 'phono-symbolic' (Malkiel 1990), 'phono-iconic' (Voronin, 2006) and 'sound symbolic' words (Hinton et al 1994), or 'expressives' (Carr 1966; Klamer 2001).

The *Oxford English Dictionary* (further abbreviated OED) interchangeably applies terms 'onomatopoeic', 'expressive', 'imitative', and 'echoic' for such lexemes. Consider, for example, the following: *bing* 'echoic, representing a sudden banging noise or blow' (OED); *blubber* 'probably onomatopoeic, from the action of the lips in making a bubble' (OED); *fuzz* 'imitative of the action of blowing away light particles' (OED); *giggle* 'echoic'; *glug* 'a word formed to imitate an inarticulate sound; echoic' (OED); *growl* 'probably an echoic formation'; *bangarang* 'probably an imitative or expressive formation' (OED). Such inconsistency in the use of terms may be due to the fact that the dictionary has been compiled and reviewed by different lexicographers during a time span of over 100 years.

Also, the term ‘ideophone’ is quite often applied to imitative words (Voeltz et al. 2001). Originally, the term was coined to describe a certain class of words in Bantu. Doke (1935, 118) described an ideophone as ‘a vivid representation of an idea in a sound; a word, often onomatopoeic, which describes a predicate, qualificative or adverb in respect to manner, color, sound, smell, action, state or intensity’. Vague as it is, this term was soon extended to iconic words of languages outside Africa, for example, English (Smithers 1954).

The terminological difficulties increase when it comes to further classification of imitative words. Hinton et al. (1994, 1-12) proposed a classification of *sound symbolic* words. They distinguish *corporeal* sound symbolism (words denoting ‘symptomatic’ sounds like hiccupping and coughing); *imitative* sound symbolism (onomatopoeic words); *synesthetic* sound symbolism (acoustic symbolization of non-acoustic phenomena) and *conventional* sound symbolism (analogical association of certain phonemes and clusters with certain meanings). The latter phenomenon is better known under the term ‘*phonaesthemes’* (see Marchand 1959)—consider, for example, English ***gl**ow, **gl**itter, **gl**isten* etc.

In this chapter I use the terminology and classification introduced by Stanislav Voronin (1969). He proposed to divide all imitative/phono-iconic (“звукоизобразительные”) words into two subclasses—onomatopoeic (“звукоподражательные”) and sound symbolic (“звукосимволические”) words. The distinction is based on the distinction of their denotata.

A denotatum of an ***onomatopoeic*** word is always a sound (e.g. an abrupt sound for English *plop* ‘an abrupt hollow sound, as of a small solid object dropping into water without a splash’ (OED); or a sibilant sound for Icelandic *ussa* [ys:a] ‘to hiss’).

A denotatum of a ***sound symbolic*** word is not a sound but an instance of movement/articulation. It could be related to physiological processes (like articulation of putting the lips together and simultaneously exhaling as in the English *puff* ‘to blow out air in a short abrupt blast’ (OED), or an articulatory gesture imitating coughing as in English *heck* ‘to cough slightly’ (OED)), or a purely imitative articulation (like articulatory imitation of a round shape of an object as in English *bubble* ‘a body of gas present in a liquid’ (OED)).

Voronin classifies onomatopoeic words further by dividing them into five (sub)classes according to their denotata (Voronin 2006, 44). He distinguishes the following ***types of sound denotata***—(1) pulses; (2) non-pulses; (3) dissonances (rapid series of pulses), stating that two of these classes are diametrically opposite (pulses and non-pulses), whereas the third one (dissonances) is in the 'middle' between the first two (Voronin 2006, 45). He also added two types of denotata combinations which he called 'hyper classes'—(4) pulse-non-pulses and (5) dissonances-quasi-pulse-non-pulses.

The three classes of onomatopoeic words which reflect these three classes of sound denotata Voronin called correspondingly: *instants*, *continuants* and *frequentatives*, and the two hyper-classes of onomatopoeic words which render two hyper classes of sound denotata—*instants-continuants* and *frequentatives-quasi-instants-continuants* respectively (Voronin 2006, 45). See Fig. 7-1.

An ***instant***, according to Voronin (2006, 45), denotes a pulse-like sound, that is, a sound which is an instantaneous (or a 'super-short') noise or tone which is perceived as an acoustic 'hit' by a human ear. In natural languages such 'hits' are typically rendered via plosives—for example, English *tap* 'to strike lightly, but clearly and audibly', Indonesian *tuk* 'a knocking sound', Russian *тук-тук* (tuk-tuk) 'knock-knock'.

Continuants is a class of onomatopoeic words denoting non-pulse natural sounds, that is, sounds of 'prolonged' and 'coherent' duration which are not divided into shorter segments. Such sound denotata are perceived either as tonal or noise phonations.

Tone continuants (Voronin 2006, 46) render tonal non-pulse sounds or tones in their purest form via vowels of different quality (the pitch of the vowel iconically rendering the pitch of the denoted sound). Compare the following denotations of howling: English *howl* (Middle English [hu:l]), *toot*; Even *u-u-u*; Selkup *lūt*; Enets *θuð* or *θūð*; Nanai *tu-u* (examples taken from Voronin 2006, 46).

Noise continuants (Voronin 2006, 46) denote noise-like non-pulse sounds, in other words, a noise in its purest, often via sibilants: English *hiss*, Turkish *fiš*-; Japanese *shü-shü,* Ossetian *syf-syf* (examples taken from Voronin 2006, 46-47).

Frequentatives is a class of onomatopoeic words denoting a rapid series of pulses where each pulse is hardly perceived separately yet there is no complete fusion of pulses into one tone; such rapid sequence of pulses is highly irritative for acoustic perception (Voronin 2006: 47). Frequentatives are most typically rendered via the phoneme /r/: English: *br-r-r* (occasional imitation of the sounds made by a drill); Bashkir *pyr(r)* 'a rumbling noise made by a car motor', Chuvash *tyrr* 'denotation of insects' chirring' (examples taken from Voronin 2006: 47-48).

It should be noted that the majority of onomatopoeic words in natural languages belong to 'mixed' hyper-classes (statistics for the English language—see Flaksman 2015: 67).

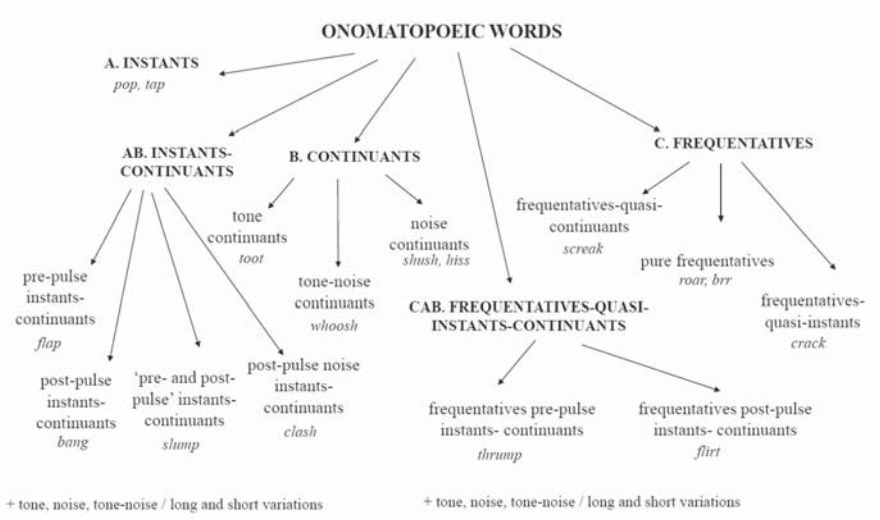


Fig. 7-1: Voronin's classification of onomatopoeic words with English examples (quoted from Flaksman 2020b, 140).

Voronin also proposed a sub-division of sound symbolic words (Voronin 2006, 70) into two categories: intrakinesemisms and extrakinesemisms.

Intrakinesemisms (e.g. E. *puff*, *heck*, *cough*, *spit*) are imitative words conveying their meaning by means of both articulation and acoustic mimicry (they are further classified into A. 'nose articulations', B. 'mouth articulations' and C. 'throat articulations'). Altogether, Voronin distinguishes 36 sub-types of intakinesemisms (see Voronin 2006, 87).

Extrakinesemisms (e.g. E. *bubble*) are purely articulatory copies; they are mimetic gestures produced by vocal cords, facial muscles and tongue; the sound in extrakinesemisms is only secondary.

3. Methods of phonosemantics and their application to the material of extinct languages

One of the main challenges of iconicity studies is establishing the imitative origin of a word. As imitative words evolve over time, their original, iconic nature becomes obscured by later—semantic and phonetic—changes. There are two methods for dealing with this problem in phonosemantics. The first is the method of phonosemantic analysis, and the second is the method of diachronic evaluation of imitative lexicon. Both methods rely heavily on the results of historical-comparative research and etymological investigations.

3.1 Method of phonosemantic analysis

3.1.1 Method description

The method of phonosemantic analysis is a method for establishing the imitative origin of a word (Voronin 1990). The method includes several steps:

I. Semantics;
II. Search for salient iconic traits;
III. Etymology;
IV. Typology;
V. Search for nomination motif;
VI. Summa summarum.

Semantics is an operation of establishing the original meaning of a word. Some words have meanings obviously related to sound (E. *buzz* ‘a sibilant hum, such as is made by bees, flies, and other winged insects’ (OED)) or to performing imitative gestures (E. *heck* ‘to cough slightly; to imitate the noise of a cough’ (OED)). However, a meaning suggestive of sound or imitation of a sound gesture might not lie on the surface—e.g. E. *fluff* ‘light, feathery, flocculent stuff’ (OED), which is a word originally imitating ‘the action of puffing away some light substance’ (OED), cf. Russian *пух* (pukh) ‘fluff’ with original meaning ‘breath, breathing’ (Vasmer 2009, vol. 3, 414).

Salient iconic traits are such traits as (1) expressive ablaut (E. *ding-dong*); (2) violation of phonotactic constraints of a language (E. ***vr****oom* ‘the roaring noise of a motor vehicle accelerating or travelling at speed’ (OED), ***thw****ack* ‘to beat or strike vigorously’ (OED)); (3) expressive vowel or consonant lengthening (occasional English *FFF FFFT* ‘the sound of an airborne object in motion’ (Taylor 2007, 51), occasional English *OOOSSH* ‘the sound of a rocket in flight’ (Taylor 2007, 103)); 4) partial or full reduplication (E. *cuckoo* ‘a bird, *Cuculus canorus*’ (OED)). The presence of these traits indirectly suggests possible imitative origin of a word.

Etymology is an essential operation of phonosemantic analysis (Voronin 2006, 88). Without a thorough, deep-down etymological analysis, a good part of imitative words of the lexicon is left out. For example, etymological analysis of English *buffoon* 'a low jester' (OED) shows that it goes down to Italian *buffare* 'to puff, as a comic gesture' (OED), which is of echoic origin (HARP). However, this operation is not always applicable. For example, there are no etymologies for invented or occasional imitative words (see Davydova 2016).

Typology is an operation of establishing a cross-linguistic similarity of imitative words. One shall bear in mind that the compared words should not be identical (which is rarely possible statistically, although there is a counterexample from section 2 – Indonesian *tuk* / Russian *тук-тук* (tuk-tuk)). The analyzed words should be compared according to their classes and hyper-classes (see section 2): for example, English *ahchoo* and Russian *апчхи* (*apchkhi*) are both sound symbolisms intrakinesemisms which share the same meaning—'to sneeze'. The idea is to prove that the iconic form-meaning correspondence in question is recurrent in various non-related languages, thus adding objectivity to the suspected form-meaning link.

Search for nomination motif also implies a detailed categorization of an imitative word. If there is no clear nomination motif, there is no iconicity. This operation answers two questions: (1) what is imitated and (2) how exactly is it imitated? For example, E. *puff* imitates the action of blowing and it has a labial plosive coupled with a labial vowel on the onset (iconic for the lip movement), labial fricative rendering the notion of exhaling. Thus, according to Voronin's classification, *puff* is a sound symbolic word (articulatory imitation), intrakinesemism (it imitates 'inner' processes of human physiological universe), type B. 'mouth articulation'. For the types of nomination motifs, see section 2.

Summa summarum—the conclusion about whether the word is imitative by origin should only be reached upon considering the sum of evidence. For instance, salient iconic traits might not be present in all imitative words, as they are more characteristic of recently coined or occasional onomatopoeias (Flaksman 2015, 174).

3.1.2 Method application

Let us consider the application of the method of phonosemantic analysis on the example of the Present-day English *clang*:

I + the word has a meaning related to sound – 'a loud resonant ringing sound' (OED);

II + it exhibits salient iconic traits—it has imitative variants *clank, clonk* with expressive vowel and consonant alternation (these are *not* the instances of IE ablaut);

III + etymologically it is traced down to PIE **klang*-, nasalized form of root **kleg*- 'to cry, sound' (HARP), so its original (reconstructed) meaning was initially related to sound;

IV ? the word is typologically similar to Viet. *leng keng* 'jingle';

V + it contains plosives for abrupt onset, and nasals for resonant sound following the hit (acoustic imitation);

VI *conclusion:* it is an onomatopoeic word, hyper-class instants-continuants.

Here is another example of the application of the method of phonosemantic analysis—English *buzz*:

I + the word has a meaning related to sound—'a sibilant hum, such as is made by bees, flies, and other winged insects' (OED);

II + it exhibits no salient iconic traits;

III + it was first recorded in the late fourteenth century English; 'from the sound' (OED); the meaning is deduced from the contexts of first quotations, e.g. 1530 J. Palsgrave Lesclarcissement 473/1 *Harke how this fleshe flye* <u>*busseth*</u>. (OED); 1556 J. Heywood Spider & Flie lvii. 241 *As if ten millions of flies had ben* <u>*buzzing*</u>. (OED);

IV + it is similar to the following words with meanings related to buzzing: E. *chizz, huzz, tizzy, whizz,* Slovak *bzzz* (Iconicity Atlas 2019) [these words contain voiced sibilants which iconically reflect the sound made by insects];

V + voiced sibilant iconically renders continuant dissonant noises made by insects;

VI *conclusion*: it is an onomatopoeic word, hyper-class instants-continuants.

Another (negative) example of the method application is that of English *mouse*:

I – the word has a meaning not related to sound—'any of numerous small rodents of the family *Muridae*, which usually have a pointed snout, relatively large eyes and ears, and a long tail, and typically feed on seeds and fruit' (OED);

II – it exhibits no salient iconic traits;
III – etymologically it is traced down to PIE **mus*- 'mouse' (HARP), so its original (reconstructed) meaning was not related to sound/articulatory gesture;
IV – the word has no phonetic similarity across non-related languages (cf. Estonian *hiir*, Turkish *fare*, Vietnamese *chuột*);
V – there is no iconic similarity between the form of the word (PIE **mus*-) and its meaning (neither with related, 'attributive' meanings, e.g. the characteristic cry of an animal);
VI *conclusion:* it is not an imitative word.

The method was successfully applied to the material of Georgian (Kankia 1988), Bashkir (Lapkina 1979), Indonesian (Bratus 1976), Estonian (Veldi 1988), and Lesgin (Mazanaev 1985). Svetlana Klimova (1986) with the method's help established the imitative origin of a number of words marked as 'of uncertain origin' in *The Concise Oxford English Dictionary*.

3.1.3 Phonosemantic analysis of extinct or reconstructed words

We have applied the method of phonosemantic analysis to words from ancient and reconstructed languages. It yielded the following results (on the example of OE *drán*):

I + the word has a meaning related to sound—'a drone [a buzzing insect]' (BT);
II – it exhibits no salient iconic traits;
III + it is related to OS *drano*, *dran,* MLG *drāne*, *drōne*; G. (regional) *drāne*, *drōne* (G. *Drohne*) and probably also to OS *dreno*, OHG *treno*, *tren*; 'probably ultimately from a Germanic verbal base for making a kind of loud, continuous sound; the noun was apparently formed from this verbal base with reference to the loud buzzing sound made by bees and similar insects, perhaps sometimes specifically with reference to the males of some species buzzing aggressively when the hive is disturbed' (OED); ultimately from Proto-Germanic **dran*- (source also of MDu *drane*; OHG *treno*; G. *Drohne*, which is from MLG *drone*), probably imitative (compare Lithuanian *tranni*, Greek *thronax* 'a drone') (HARP).

IV ?? might be similar to Lithuanian *trãnas* ‘a drone’, Latvian *trans* ‘a drone’, Russ. *трутень* ‘a drone’ (HARP; also Vasmer 2009, vol. 4, 111); with more distant similarity—cf. Russ. *трещать* ‘to chirr, crackle’;
V + /r/ corresponds to a low vibrant sound made by the insect;
VI *conclusion*: it is an onomatopoeic word, class frequentatives-instants-continuants.

An example of phonosemantic analysis of a reconstructed word is Proto-Germanic **fneusan- / *fnūsan-* ‘to sneeze’ (Kroonen 2013, 149):

I + the word has a meaning related to an articulatory gesture—‘to sneeze’;
II – it exhibits no salient iconic traits;
III + Proto-Germanic **fneusan- / *fnūsan-* ‘to sneeze’—ON *hnjósa*, Far. *njósa*, Nw. *njosa*, *nysa*, OE *fneosan*, ME *fnesen*, E. *sneeze*, MDu *niesen, fniesen*, Du *niezen*, obs. *fniezen*, MHG *pfnūsen* ‘to snort, sneeze’; from IE **pneus-e-* (cf. Gr. *πνέω* ‘to blow, cough, breathe’), ‘ultimately, the root is likely to be onomatopoetic’ (Kroonen 2013, 149);
IV ? G. *fnehan* ‘to snort, breathe heavily’ (Kroonen 2013, 149);
V + /n/ corresponds to a ‘nasal gesture’, /p >f/—to lip movements;
VI *conclusion*: it is sound-symbolic word, intrakinesemism, class A. ‘nose articulations’.

The application of the method of phonosemantic analysis makes it possible to establish the imitative origin of ancient and reconstructed words, however, it has several limitations—the older the word is, the move difficulties there are in establishing its imitative origin:

- one has to rely heavily on the effort of previous research, especially when it comes to reconstructed forms and meanings;
- the original meaning becomes ‘blurred’—the number of contexts indicating the word’s meaning in ancient languages is limited, and in the reconstructed the context is absent;
- salient iconic traits (if the word exhibited such) may not be evident from the recordings or reconstruction (English

barbarous, barbarian are the exceptions where reduplication—a salient iconic trait—can still be traced);
- parts of compound words as well as morphological affixes can blend, thus hindering the reconstruction (e.g. English *rhubarb*—via French, from post-classical Latin *reubarbarum*, probably from Hellenistic Greek *ῥῆον βάρβαρον* (OED)—for discussion on the point see (Flaksman 2020a));
- the form of old imitative roots may have already changed several times by the time of their recording; thus, tracing down the words' nomination motif might present some difficulties;
- upon typological comparison one should bear in mind that 'old' imitative words conformed to phonological and phonotactic rules that were at work at the time of their coinage; whereas 'young' (especially recently coined or occasional) imitative words may violate these rules (see examples *vroom* and *thwack* above).

3.2 The method of diachronic evaluation of imitative lexicon

The second method is the method of diachronic evaluation of imitative lexicon. It helps to differentiate, on a synchronous level, words on different stages of de-iconization.

3.2.1 Method description

De-iconization is a gradual loss of iconicity caused by parallel acting of regular sound changes and regular sense development of a word (Flaksman 2015, 120; see also Flaksman 2017).

Diachronic development inevitably tears apart the original sound-meaning link of an imitative word, thus lessening its expressivity and obscuring its iconic origin. As a result, on a synchronous level there co-exist words on different stages of de-iconization. Altogether, I distinguish four stages of de-iconization, from the least de-iconized (SD-1) to the most de-iconized (SD-4) (Flaksman 2015, 126). To distinguish the de-iconization stages one has to evaluate a word according to several parameters. Thus, the method includes three steps:

I. establishing the word's degree of phonetic, morphological, and syntactic integration—newly coined imitative interjections like *zzz, brrr* poorly fit into a language and exhibit salient iconic traits;

II. establishing the regular sound changes a word has undergone and whether these sound changes were phonosemantically significant;
III. establishing semantic shifts the word has undergone and whether it still retains its original (sound-related) meaning.

The division of imitative words into SDs can be summarized as follows (Table 7-1):

	SD-1	**SD-2**	**SD-3a**	**SD-3b**	**SD-4**
integration	–	+	+	+	+
regular sound changes	–	–	+	–	+
considerable semantic shifts	–	–	–	+	+

Table 7-1: Stages of de-iconization of imitative words.

The most vivid, newly coined imitative words (words classified as ***SD-1***) are, at the same time, the least conventional and thus, poorly integrated into the language both syntactically and morphologically. These are interjections, or in some languages, ideophones. For example, see the occasional English *FFF FFFT* ‘the sound of an airborne object in motion’ (Taylor 2007, 51), discussed in section 2. Consider also *ahchoo* ‘imitative of the sound made when sneezing’ (OED); *bam* ‘an interjection imitating the sound of a hard blow’ (OED); *boing* ‘the noise made when a compressed spring is suddenly released; a reverberating sound’ (OED); *bom* ‘the sound caused by the discharge of a gun, less deep and sonorous than a ‘boom’ (OED).

Words classified as ***SD-2*** are nouns, verbs, and adjectives. They are well-integrated and expressive at the same time. Iconicity (imitativeness) is lessened by (conventional) suffixes and inflectional morphology (Russ. *кудахт-а-ть* (kudakht-a-t’) ‘to cackle’; German *gluck-en* ‘to cackle’; Icelandic *gagg-a* ‘to cackle’), see also (McGregor 2001). Words classified as SD-2 have less freedom in form variation, yet the number of the contexts they can be used in increases. Cf. English *ahchoo* (SD-1) and *sneeze*—(SD-2)—1986 S. Gray Sex & Death to Age 14 71 *Ah, ah, ah, ahchew! Oh, Melanie look! I sneezed my brains out!* (OED); also [occasional] English *ptooey* (SD-1) and *spit* (SD-2).

Words classified as ***SD-3*** have undergone either form or meaning changes, thus, having their expressivity lessened. Consider, for example, an SD-3a word *chirr* /tʃəː/ ‘to make the trilled sound characteristic of grasshoppers' (OED), which after the vocalization of /r/ in British English lost its 'core' element /r / 'responsible' for imitation. Regular sound changes considerably lessen original form-meaning correlation of imitative words. However, they are not the leading factor of de-iconization (Flaksman 2015, 153). Semantic shifts destroy iconicity more effectively and more frequently.

Words classified as ***SD-4*** can only be named 'iconic' from the historical perspective. For example, English *barbarous* /ˈbɑːbərəs/ 'foreign, outlandish; uncultured, uncivilized' comes down to Latin *barbarous*, which is from Greek *βάρβαρος*; the Greek word had probably a primary reference to speech; the sense-development in ancient times was (with the Greeks) 'foreign, non-Hellenic,' later 'outlandish, rude, brutal'; (with the Romans) 'not Latin nor Greek,' then 'pertaining to those outside the Roman empire'; hence 'uncivilized, uncultured,' and later 'non-Christian,' whence 'Saracen, heathen'; and generally 'savage, rude, savagely cruel, inhuman' (OED). Imitative origin of *barbarous* is indirectly suggested by reduplication (**bar-bar*). However, both the the meaning of the word and its form (in the course of borrowing and under the influence of regular sound changes) and have changed beyond recognition.

3.2.2 Method application

Let us consider the application of the method of diachronic evaluation using the example of Present-day English *fnar* (also *fnarr, fnar fnar*), a sound symbolic word-intrakinesemism, class A. 'nose articulations' (examples taken from the corresponding entry in the OED):

I. a. the word is registered as an interjection (only occasionally as an adjective):

2003 G. Burn *North of Eng. Home Service* (2004) iii. 80 His friend sniggered, and the boy who had told the joke, exceedingly red now, sniggered as well. '*Fnarr.*';
2000 Campaign (Nexis) 5 May The campaign used obvious innuendo and lots of 'fnar fnar' jokes;

b. the word is often reduplicated; it contains an initial consonant cluster *fn*- which violates phonotactic rules of the Present-day English;

II. the word was coined in 1987 in Britain (OED); no regular sound changes since;
III. the word has the only recorded meaning suggestive of an articulatory gesture: ‘lecherous or half-suppressed laughter’.

Thus, *fnar* is an SD-1 word—it exhibits poor phonetical and morphological integration; it has not yet undergone any drastic form or meaning changes. Therefore, it is the most iconic (SD-1).

Let us consider English *clang*:

I. in Present-day English the word is both noun and verb (OED); it is morphologically and syntactically integrated; imitative variants (*clank, clonk)* indicate traces of (earlier?) form variation;
II. it is either coined in the late sixteenth century (first recorded 1576 (OED)) or borrowed from Latin (OED), no regular sound changes since;
III. the word retains its original sound-related meaning—*a clang* ‘a loud resonant ringing sound’; *to clang* ‘to emit a loud resonant ringing sound as of pieces of metal struck together’ (OED).

Thus, English *clang* is currently classified as SD-2—it is integrated but not yet transformed.

The third example of the method application is that of *bop* (US, colloquial) ‘to hit, strike, punch, to fight’ (an onomatopoeic word, an instant).

I. it is a verb—see the quotation: 1938 P. G. Wodehouse *Summer Moonshine* xv. 181 I’ll <u>bop</u> you over the head with this chair (OED);
II. the word was coined in the late nineteenth century; it has undergone no regular sound changes;
III. the word retains no meaning related to sound in current usage; it has undergone the metonymic shift [‘an abrupt sound accompanying the hit’] > ‘to hit’ (the sound-related meaning was registered in early dialectal use—W. D. Parish & W. F. Shaw Dict. Kentish Dial. 16 <u>*Bop*</u>, to throw anything down with a resounding noise).

Thus, English *bop* 'to hit' is a SD-3b word—it is integrated, it retains its original form, but has lost its original meaning. The iconic sound-meaning link is 'broken' by a semantic shift.

3.2.3 Diachronic evaluation of extinct or reconstructed imitative lexicons

The method of diachronic evaluation of imitative lexicon can be successfully applied to the material of ancient and reconstructed languages. Let us consider Old English *cracian* 'to crack, sound, resound' (an onomatopoeic word, a frequentative-instant):

I. it is a verb, syntactically integrated into a sentence:

   Ps. Th. 45, 3. *Sió eorþe eall cracode*. 'The whole earth quaked' (BT);

II. *cracian*, Present-day English *crack* goes back to Proto-Germanic **krakojan* (source also of Middle Dutch *craken*, Dutch *kraken*, German *krachen* (HARP)); no considerable regular sound changes between Proto-Germanic and Old English;

III. the word in Old English denotes a sound—'to crack, sound, resound' (BT).

Thus, *cracian* is an SD-2 word—it was integrated in Old English and had not undergone any significant phonetic or semantic changes.

Let us also consider Old English *drán* 'a drone':

I. it is a noun, syntactically integrated into a sentence:

   Chr. 1127; Erl. 256, 20, 21. *Ðǽr he wunede eall riht swá dráne dóþ on híue: eall ðæt ða beón dragen toward ða dráne dragaþ fraward he abode*. 'There just as drones do in a hive: all that the bees draw towards them the drones draw from them.' (BT);

II. the word could be traced at least to Proto-Germanic **dran-* with no considerable regular sound changes between Proto-Germanic and Old English;

III. the word has undergone a semantic shift [a low-pitched vibrant monotonous sound] > an insect [producing this sound].

Thus, *drán* is an SD-3b word—it was integrated in Old English and had not undergone any significant phonetic changes, but had undergone a semantic shift.

Here is an example of a reconstructed word—Proto-Germanic **fīsan-* 'to blow' (Kroonen 2013: 142):

I. it is reconstructed as verb;
II. it probably comes down to PIE **psei-s-e-* (PIE form of the verb may have been **psei-s-e-* (Kroonen 2013: 142)); then there is the *p* > *f* change (Grimm's law), although both articulations are labial (therefore, the change is not phonosemantically significant);
III. no drastic semantic shifts in PGerm, although later 'the semantic spectrum [of this originally onomatopoetic verb] ranges between 'to blow', 'to hiss', 'to whisper' and 'to fart' (Kroonen 2013, 142).

Upon this evidence, we might suggest SD-2 inclining towards SD-3b for the Proto-Germanic **fīsan-*.

The application of the method of diachronic evaluation makes it possible to distinguish imitative words belonging to different de-iconization stages on all synchrony levels. However, it should be noted that application of the method to ancient and reconstructed languages has certain limitations:

- the synchrony of ancient and reconstructed languages is rather conventional—for instance, Old English encompasses the timespan of about four centuries, the words might have changed their de-iconization stage within this period without the researchers being able to trace the change on the basis of the recorded texts;
- the lack of or a limited context can give only approximate judgement about the word's de-iconization stage as the word might have had a non-recorded sound-related meaning;
- it is impossible to find non-integrated SD-1 words in reconstructed languages;

- SD-1 words are extremely rare in ancient and extinct languages (due to the nature of the texts—for a discussion, see Flaksman 2019);
- words classified as SD-4 could be found in languages with a thoroughly researched predecessor (thus, it is theoretically possible to find SD-4 words in Old English and even Proto-Germanic, whereas they would be indistinguishable in PIE);
- SD-3b and SD-4 words in ancient and reconstructed languages could be identified mostly by using the method of semantic reconstruction, e.g. the bird names can be ascribed onomatopoeic origin upon the common semantic shift 'imitation of the call' > 'the name of the bird').

4. A study of imitative words in Holthausen's *Altenglishes Etymologishes Wörterbuch*

The two methods discussed above were applied to the material of the Old English language (also see Flaksman 2019, 1072-1073; Flaksman 2021 [in print])). The ***material*** for this purpose was compiled from the list of imitative (onomatopoeic and sound symbolic) words obtained by the method of continuous sampling from the Holthausen's *Altenglishes Etymologishes Wörterbuch* (1974).

Some entries of this etymological dictionary contained the direct indication of the word's imitative origin. For example, *hi*! is labelled 'Schallwort', that is 'onomatopoeic word', in the dictionary (Holthausen 1974,158). Other iconic words were identified by authors of other etymology dictionaries (BT; OED; HARP; Kroonen 2013; Skeat 2006; Onions 2002; Barnhard 2006; Ásgeir Blöndal Magnússon 1989; Kluge 2011; Lehmann 1986; Feist 1888).

For example, *fnéosan* 'to sneeze' (Holthausen 1974, 110) comes down to Proto-Germanic **fneusan-* which 'is likely to be onomatopoeic' (Kroonen 2013, 149); OE *bealcan* 'to belch' is also not marked 'imitative' in Holthausen (1974, 110), but marked 'of echoic origin' in (HARP); the same is applicable to OE *cnyllan* 'to knell' (Holthausen 1974, 55) which is marked 'imitative' in (HARP).

Also, the dictionary contained a number of words iconic origin of which was established by the method of phonosemantic analysis (see section 3.2.3 and the example of *cracian*).

Altogether, 300 imitative words were identified in the dictionary. For example: *breahtm, bearhtm, beorhtm, brehtm*, 'a noise, tumult, sound, cry'—Exon. 36 a; Th. 118, 1; Gú. 233. *Þá wearþ* <u>*breahtm*</u> *hæfen*. 'Then a noise was raised' (BT); *cnocian, cnucian, cnucigan* 'to knock, beat, pound'—Homl. Th. ii. 382, 17, 22. *He cnucode æt ðære dura.* 'He knocked at the door.'(BT); *hleahtor, hlehter* 'laughter'—Beo. Th. 6033; Se herewísa hleahtor álegde. 'The host's leader hath put away laughter [is dead].' (BT); *hlin, hlyn, hlynn* 'a sound, noise, clamour, din'—Cd. 119 *Hlynn wearþ on ceastrum* 'A great cry arose in the cities'(BT); *sweg* 'unregulated, confused sound, noise, din, crash; sound made by living creatures, voice, cry or note of a bird, song; sound made by means of an instrument'—Ælfc. Gr. 5 *Sume synd geworhte æfter gelícnysse ágenes swéges,* turtur *turtle*, (BT); 'Some are made after the likeness to their own sounding: *turtur* 'turtledove'. Further examples and discussion – see Flaksman 2021 [in print].

The application of the method of diachronic evaluation of imitative lexicon has shown that SD-2 words make up more than a half of the analyzed material (169, or 56%), SD-3 words constitute 35% of the corpus (104 words), words classified as SD-1 and SD-4 are quite rare (6, or 2%) and (21, or 7%) respectively (see also Flaksman 2019):

SD-1	*ea*! *eáw*!	'alas!'
SD-2	*hlynn*	'a sound, noise, clamour, din'
	sweg	'sound, noise, din, crash'
	hleahtor	'laughter'
	cnocian	'knock'
	brehtm	'a noise, tumult, sound, cry'
SD-3	*galdor*	'incantation'
	turtle	'turtle-dove'
SD-4	(?) *su*	'a sow'

It should be noted that the majority of the studied Old English imitative words do not survive until today – only 112 (37%) of the words are preserved in Present-day English. But if we have a closer look at those 112 words we will see that 40% of them (45 words) have changed their de-iconization stage. For example, OE *beorcan* (SD-2) is Present-day (British) English *bark* [ba:k] (SD-3a)—because of the /r/ vocalization (16-17th century) it has changed its form (thus weakening the iconic form-meaning link).

On the whole, imitative words coined before Middle English differ from the later ones—two languages limit the means of imitation by their differences in the phonemic inventory:

- Old English had geminates *t:, d:, p:, b:, k:, k':, g:* (e.g. *clæ**pp**an*);
- Old English had phonemes *y* [ü] (e.g. *d**y**nnan*); *χ, γ*, and palatal consonants;
- in Old English *v, z, ð* were voiced allophones of *f, s, θ,* whereas in Present-day English they are phonemes and could be placed initially (English *zoom*),
- English initially did not have phonemes *tʃ, ʤ, ʃ, ʒ* (e.g. E *mun**ch***);
- Old English had now prohibited phoneme sequences *hl-, hw-, hn-, hr-, kn-, gn-* (***gn**ornian, **cn**ocian*);
- Present-day (British) English prohibits a post-vocal /r/, whereas it was normal in Old English (*ra**r**ian, beo**r**can*).

All these changes in the phonemic inventory and in the rules of phonotactics have affected its imitative vocabulary.

Regular sound changes have led to the de-iconization of certain classes of onomatopoeic words (namely frequentatives and frequentatives-instants-continuants), on the other hand, they gave room for the new imitative coinage. For instance, *sk > ʃ* change in the ninth century led to appearance of the phoneme /ʃ/ now used in imitative coinage: *shoo* (1483) 'an exclamation used to frighten or drive away poultry, birds, or other intruders'; *shush* (1924) 'to call or reduce (someone) to silence' (OED).

Here one should ask oneself this important question—if a language constantly changes structurally (which predetermines de-iconization of imitative words) and also is renewing its vocabulary (two-thirds of Old English imitative words are now obsolete), are imitative systems on different synchronic levels of the language's existence one and the same or different? I am inclined to answer this question negatively. At least, when the English language is concerned. These are my arguments:

- more than a half of the Old English imitative words have disappeared physically over 1000 years;
- Old and Present-day English imitative words differ structurally (on the whole, they contain different phonemes, and these phonemes are combined differently due to the changed structural (conventional) restrictions);
- 40% of the surviving Old English imitative words have changed their de-iconization stage (SD-2 > SD-3a, SD-2 > SD-3b, SD-3a > SD-4 etc.);
- only 27 words remained in the core of the language's imitative lexicon (that is they are still classified as SD-2); these are *ca* 10% of the studied 300 Old English lexemes (the discussion about

what is a core, and what is a periphery of the imitative lexicon—see Flaksman 2019).

Thus, the study of the Old English imitative lexicon conducted on the material of Holthausen's *Altenglishes Etymologishes Wörterbuch* (1974) has confirmed that the methods of phonosemantic analysis and diachronic evaluation of imitative lexicon can be applied to the material of an extinct language. The results of the study show dramatic differences in Old and Present-day English imitative systems.

5. Conclusions

The research conducted for this project has revealed that etymological dictionaries provide information on imitative lexicons during different stages of a language's history. Furthermore, it has shown that an imitative lexicon differs considerably in diachrony and that the lexicon's core changes dramatically over time—old imitative words become obsolete and are replaced by newer, more expressive, coinages, whereas the old, still remaining ones gradually lose an iconic link between form and meaning due to the internal factors of language evolution—regular sound changes and semantic shifts, which gradually turn iconic words into 'ordinary' ones with an arbitrary sound-meaning correlation.

Abbreviations

BT – *An Anglo-Saxon Dictionary*, ed. by J. Bosworth and T.N. Toller
Du – Dutch
E. – English
Far. – Faroese
G. – German
Gr. – Greek
HARP – *Online Etymology Dictionary*, ed. by D. Harper
MDu – Middle Dutch
ME – Middle English
MHG – Middle High German
MLG – Middle Low German
Nw. – Norwegian
OE – Old English
OED – The Oxford English Dictionary, 3d [online] edition (consulted between June 2019 and March 2020)
OHG – Old High German

OS – Old Saxon
PGerm – Proto-Germanic
PIE – Proto-Indo-European
Russ. – Russian
SD – stage of de-iconization

Bibliography

Afanasiev, Andrei Jurievich. 1984. "Voprosy Semanticheskoi Evolutsii Leksiki: Na Materiale Angliiskikh Zvukopodrazhatelnikh Suschestvitelnikh [Aspects of Vocabulary's Semantic Evolution: A Study of English Imitative Words]". PhD diss., University of Leningrad.

Anderson, Earl. 1998. *Grammar of Iconism*. London: Associated University Press.

Ásgeir Blöndal Magnússon. 1989. *Íslensk orðsifjabók*. Reykjavík: Orðabók Háskólans.

Barnhart, Robert K, ed. 2006. *Chambers Dictionary of Etymology*. New York: Chambers.

Bosworth, Joseph, et al. 2013. "An Anglo-Saxon Dictionary Online." Edited by Thomas Northcote Toller and Others, Faculty of Arts, Charles University in Prague, 27 Oct. 2013, Accessed 9 Jan. 2020. https://bosworth.ff.cuni.cz/018075.

Bratus, Igor Borisovich. 1976. "*Akusticheskie Onomatopi v Indoneziiiskom Jazike* [Acousic Onomatopoes in Indonesian]". PhD diss., University of Leningrad.

Carling, Gerd, and Johansson, Nilkas. 2014. "Motivated language change: processes involved in the growth and conventionalization of onomatopoeia and sound symbolism." *Acta Linguistica Hafniensia*, 46 (2): 199-217.

Carr, Denzel. 1966. "Homorganicity in Malay-Indonesian in Expressives and Quasi Expressives." *Language. Journal of the Linguistic Society of America*, 42 (2): 370-377.

Davydova, Varvara Alexeevna. 2016. "Sound Symbolism in Invented Languages". In *Anglistics of the XXI Century*, vol. 2. *Phonosemantics*, edited by Maria A. Flaksman and Olga I. Brodovich., 32-39. St. Petersburg: ANCO 'University Education Districts".

Dingemanse, Mark. 2012. "Advances in the Cross-Linguistic Study of Ideophones". *Language and Linguistics Compass,* 6 (10): 654-672.

Doke, Clement M. 1935. *Bantu Linguistic Terminology.* London: Longmans.

Feist, Sigmund, 1888. *Grundriss der Gotischen Etymologie*. Strassbourg: Trübner.

Flaksman, Maria A. 2015. "Diachronicheskoye Razvitie Zvuloizobrazitelnoi Leksiki Angliiskogo Yazyka [Diachronic Development of English Iconic Vocabulary]". PhD diss., University of St. Petersburg.

Flaksman, Maria A. 2019. "From IE *ue- to English Window: On the Number and Age of Imitative Words in the Language". In *Indo-European Linguistics and Classical Philology XXIII* (Joseph M. Tronsky memorial Conference), edited by Nikolai N. Kazansky, 1066-1075. St. Petersburg: Nauka.

Flaksman, Maria A. 2017. "Iconic Treadmill Hypothesis. The Reasons Behind Continuous Onomatopoeic Coinage". In *Dimensions of Iconicity* [Iconicity in Language and Literature 15], edited by M. Bauer, A. Zirker, O. Fischer and Ch. Ljungberg, 15-38. Amsterdam: John Benjamins.

Flaksman, Maria A. 2020a. "Pathways of De-Iconization: How Borrowing, Semantic Evolution, and Regular Sound Changes Obscure Iconicity". In *Operationalizing Iconicity* [Iconicity in Language and Literature 17], edited by Perniss, Pamela, Olga Fischer and Christina Ljungberg, 75-103. Amsterdam: John Benjamins.

Flaksman, Maria 2020b. "Stanislav Voronin's Universal Classification of Onomatopoeic Words: a Critical Approach (Part 1)". In *Discourse*, 6(4): 131-149.

Flaksman, Maria 2021 (in print). "Echoes of the past: Old English onomatopoeia". In *Iconicity in Language and Literature*. Olga Fischer and Christina Ljungberg (eds.). Amsterdam: John Benjamins Publishing Company.

Harper, Douglas. "Online Etymology Dictionary". Accessed December 21, 2019. http://www.etymoline.com.

Hinton, Leanne, Nichols, Johanna, and Ohala, John J. (eds). 1994. *Sound Symbolism*. Cambridge: Cambridge University Press.

Holthausen, Ferdinand. 1974. *Altenglishes Etymologishes Wörterbuch*. Heidelberg: Carl Winter Universitätsverlag.

Iconicity Atlas Project: A Multilingual Online Dictionary of Iconic Words and Ideophones, edited by Maria A. Flaksman and Natalia N. Noland. Accessed December 12, 2019.
http://www.iconicity-atlas.com/index.htm.

Kankia, Noe Davidovich. 1988. "Primarnaya Motivirovannost Slova (na Materiale Angliiskogo i Grizinskogo Yazikov) [The Primary Motivation of The Word (In English and Georgian)]." PhD diss., University of Leningrad.

Klamer, Marian A. F. 2001. "Expressives and Iconicity in the Lexicon." In *Ideophones. Typological Studies in Language 44*, edited by Voeltz, E. F K. and Ch. Kilian-Hatz, 165-182. Amsterdam-Philadelphia: John Benjamins Publishing Company.

Klimova, Svetlata Vladimirovna. 1986. "Glagoly 'Nejasnogo Proiskhohdenija v Sokraschennom Oxfordskom Slovare: Elementy Etimologicheskoi Fonosemantiki [Verbs of 'Uncertain Origin' in *The Concise Oxford English Dictionary*: Elements of Etymological Phonosemantics]." PhD diss., University of Leningrad.

Kluge, Friedrich. 2011. *Etymologisches Wörterbuch der deutschen Sprache*. Berlin; New York: De Gruyter.

Kroonen, Guus. 2013. *Etymological Dictionary of Proto-Germanic*. Leiden-Boston: Brill.

Lapkina, Larisa Zinnatovna. 1979. "Angliiskie I Bashkirskie Akusticheskie Onomatopi [English and Bashkir Acoustic Onomatopes]". PhD diss., University of Leningrad.

Lehmann, Winfred Philipp. 1986. *A Gothic Etymological Dictionary*. Leiden: Brill.

Malkiel, Yakov. 1990. *Diachronic Problems in Phonosymbolism*. Amsterdam: John Benjamins.

Marchand, Hans. 1959. "Phonetic Symbolism in English Word-Formation". *Indogermanische Forschungen.* 64 (2): 146-168, (3): 256–267.

Mazanaev, Ismail Abdulkadirovich. 1985. "Osnovnie Gruppy Zvukosimvolicheskih Slov: Fonosemanticheskii Analiz: na Materiale Angliiskogo i Lesginskogo Yazulov [The Main Groups of English and Lesgin Iconic Words: a Phonosemantic Analysis]". PhD diss., University of Leningrad.

McGregor, William B. 2001. "Ideophones as the Source of Verbs in Nothern Australian Languages". In *Ideophones. Typological Studies in Language 44*, edited by Voeltz, E. F K. and Ch. Kilian-Hatz, 205-222. Amsterdam-Philadelphia: John Benjamins Publishing Company.

Mithun, Marianne. 1982. "The Synchronic and Diachronic Behavior of Plops, Squeaks, Croaks, Sighs, and Moans". *International Journal of American Linguistics,* 48 (1): 49-58.

Onions, Charles Talbut. 2002. *The Oxford Dictionary of English Etymology.* Oxford: Clarendon Press.

Peirce, Charles Sanders. 1982. *Writings of Charles S. Peirce, a Chronological Edition*. Peirce Edition Project (eds.), Bloomington and Indianapolis: Indiana University Press.

Saussure, Ferdinand de. 1916. *Cours de linguistique generale*. Lausanne and Paris: Payot.

Skeat, Walter W. 2006. *An Etymological Dictionary of the English Language*. Oxford: Clarendon Press.

Smithers, G.V. 1954. “Some English Ideophones”. *Archivum Linguisticum,* 6 (2): 73-111.

Taylor, Kevin J. 2007. *KA-BOOM! A Dictionary of Comic Book Words, Symbols and Onomatopoeia*. Surrey BC: Mora Publications.

The Oxford English Dictionary (3d edition). Accessed December 21, 2019. http://www.oed.com

Vasmer, Max. 2009. *Etymologicheskii Slovar’ Russkogo Yazyka* [Etymology Dictionary of the Russian Language], in 4 volumes. Moscow: Astrel.

Veldi, Enn Antsovich 1988. “Anglo-Estonskie Paralleli v Onomatopee [English-Estonian Parallels in Onomatopoeia]”. PhD diss., University of Tartu.

Voeltz, E.F.K. and Kilian-Hatz, Christa (eds). 2001. *Ideophones. Typological Studies in Language 44*. Amsterdam-Philadelphia: John Benjamins.

Voronin, Stanislav Vasilievich. 1990. “O Metode Fonosemanticheskogo Analiza [Concerning the method of phonosemantic analysis].” In *Lingvometodicheskie aspekty semantiki i pragmatiki teksta* [Linguistic Aspects of Text Semantics and Pragmatics], 98-100. Kursk: Kursk University Press.

Voronin, Stanistav Vasilievich. 1969. “Angliiskie onomatope: Tipy i Stroenie [English Onomatopoes: Types and Structure]. PhD diss., University of Leningrad.

Voronin, Stanistav Vasilievich. 2006. *Osnovy Phonosemantiki* [The Fundamentals of Phonosemantics]. Moscow: Lenand.

Wescott, Roger W. 1980. *Sound and Sense. Linguistic Essays on Phonosemic Subjects.* Lake Bluff: Jupiter Press.

Acknowledgments: This work was funded by the Alexander von Humboldt Foundation as a part of the research project ‘Evolution of Imitative Lexicons: A Historical-Comparative Analysis of Onomatopoeia in Germanic Languages’.

CHAPTER EIGHT

SCIENTIFIC EPONYMS THROUGHOUT THE HISTORY OF ENGLISH SCHOLARLY JOURNAL ARTICLES

KATRIN MENZEL

Abstract

Eponymous expressions in academic English come in a variety of structures, e.g. proper noun + noun constructions such as *Golgi apparatus* or *Boltzmann equation*. Particularly older eponyms that were already in use in the early academic publications may have adjectival or possessive forms (e.g. *Napierian logarithm, Hadley's quadrant*). More complex term patterns, eponyms with prefixes and forms coined by minor word-formation processes are more often found in recent academic texts. Although eponyms play an important role in scientific discourse, they are an under-researched topic in terminological and lexicographical work with regard to many subject areas and scientific disciplines. Lexicographical products covering scientific eponyms contribute to standardization and documentation processes, but face a challenge to cover these forms in a systematic way, to sort them according to linguistic, chronological or thematic criteria and to keep up with ongoing developments related to their usage. It is difficult to find studies and mono- or bilingual resources on specialized eponyms in various fields, which may lead to translation difficulties when translating specialized texts into or from English. It is also hard to find specialized, searchable data that may serve to underpin assumptions and intuitions on frequencies and the productivity of eponymous terms and their underlying word formation rules. Phrasal and collocational combinations as well as possible ambiguities of eponyms are not covered sufficiently in existing resources and might be something dictionary users would like to see included in a specialized dictionary.

This article aims to demonstrate that such resources can greatly benefit from insights obtained from newly available, systematically enriched specialized electronic corpora. A recently compiled large diachronic dataset with fine-grained linguistic and metadata annotations is the Royal Society Corpus (RSC) containing around 48,000 digitized Early Modern, Late Modern and Present-Day English texts from scientific journals. It offers the possibility to trace the development of eponymous terms from the early stages of scholarly journals to contemporary publications. The results of semi-automatic extraction methods set eponymous academic terms in a broader historical framework and provide a better understanding of the development of typical word formation patterns in eponymization processes. They also shed more light on the frequencies of eponymous terms in specific disciplines. Furthermore, the results of the quantitative and qualitative analyses can be used to complement synchronic and diachronic corpus-based dictionaries as well as general and specialized dictionary databases with natural and authentic examples within their respective contexts.

1. Introduction

Personal names are a productive resource as components of new lexemes in English for specific purposes and in scientific discourse. Eponyms have played and continue to play an important role in academic discourse across disciplines. In this context, examples from medical terminology, anatomical and biological nomenclature derived from proper names may come to mind first as these are relatively well-documented fields with regard to terminology and with a long tradition to name newly discovered and recently identified or described concepts after prominent people. Other branches of science also use a number of eponyms but there are few resources providing a systematic overview of such forms in their respective fields. Individual cases of scientific eponyms used in modern languages are covered, sometimes together with other scientific terms, in specialized lexicographical resources and word-formation studies, but regularities and trends of eponymization processes throughout the history of scientific writing are relatively under-researched, even for well-documented languages such as English. One of the reasons for this research gap has been the scarcity of suitable, well-annotated and sufficiently large specialized and diachronic corpora that would allow a systematic retrieval of eponyms.

The socio-scientific and linguistic information contained in eponymous lexemes, derivatives and multi-word constructions is highly relevant for the reconstruction of various aspects of our linguistic, cultural and scientific heritage. This study aims to fill the identified research gap and to contribute to a better understanding of English eponymic processes on the basis of a diachronic dataset covering more than 330 years—the Royal Society Corpus (RSC, Kermes et al. 2016; Fischer et al. 2020).[1] The RSC is an exhaustive domain-specific corpus with full texts from selected prestigious scientific journals with numerous landmark papers that have impacted science across the globe. This makes it a unique resource that complements other existing resources in various ways. It contains around 48,000 digitized scientific journal articles from the 1660s until the 1990s with nearly 300 million tokens in total. The analysis in this article is focused on certain types of eponyms that can be queried and retrieved due to similar morphological or other structural characteristics. The results of the quantitative and qualitative analysis contribute to a better understanding of the productivity and development of typical word formation patterns in eponymization processes and of the purposes and frequencies of eponymous terms in scientific contexts they most typically occur in. They may also serve as a basis for reconstructing cross-cultural aspects of eponyms in scientific discourse from the early stages of the first scholarly journals to contemporary scientific publications. Moreover, the identified eponyms can complement synchronic and diachronic corpus-based dictionaries and lexicographical resources for specialized languages.

2. Corpus description

The Royal Society Corpus is a large diachronic specialized corpus of scientific English with full articles from the early stages of the first English scholarly journals from the transition period between Early Modern English and Late Modern English to contemporary scientific articles from Present-Day English. The RSC currently covers nearly all digitized articles from 1665-1996 from two of the world's longest-running, well-established multidisciplinary academic journals—the *Philosophical Transactions* and *Proceedings* of the Royal Society of London—and from their more specialized successor journals, the A series of the *Philosophical Transactions* and the *Proceedings* covering the mathematical and physical sciences and the B series of these journals covering the biological sciences. The *Philosophical Transactions* were split into these two thematic series A

[1] The corpus has been compiled in a project funded by the Deutsche Forschungsgemeinschaft—Project-ID 232722074—SFB 1102. The author also wishes to thank an anonymous reviewer whose constructive questions and detailed comments on an earlier draft of the manuscript led to several clarifications.

and B in 1887, and *Proceedings* in 1905. Articles from a few other more recently launched Royal Society journals on specific academic subfields that partly had been covered in the earlier *Philosophical Transactions* and *Proceedings* before are also part of the corpus. This includes texts from the *Notes and Records* journal dedicated to the history of science and the history of the Royal Society as a scientific community and biographical essays from the *Biographical Memoirs of Fellows of the Royal Society* journal with a focus on the scientific achievements of eminent Fellows or Foreign Members of the Royal Society.

In contrast to many other corpora, a large proportion of this full corpus is freely available. The Saarbrücken Collaborative Research Centre, in which the RSC has been compiled, is committed to making as much scientific content freely available as possible and providing long-term accessibility of research results and public access to research data. All corpus texts from the Late Modern English period (LModE, ca. 1700-1900) and from the transition periods at the beginning and the end of LModE have already been made available for free download and online query in a CQPweb[2] interface from the CLARIN-D centre at Saarland University under a persistent identifier (RSC 6.0 Open, ca. 17,500 texts from 1665 to 1920, ca. 78,600,000 tokens, cf. Fischer et al. 2020). The full version (RSC 6.0 Full, ca. 48,000 texts from 1665 to 1996, ca. 296,000,000 tokens, cf. Table 8-1) is available onsite to researchers and students at Saarland University.[3]

Time period	**Texts**	**Tokens**
1665–1699	1,325	2,582,856
1700–1749	1,686	3,414,795
1750–1799	1,819	6,342,489
1800–1849	2,774	9,112,274
1850–1899	6,754	36,993,412
1900–1949	10,011	65,431,384
1950–1996	23,468	172,018,539

Table 8-1: Size of Royal Society Corpus (RSC 6.0 Full).

[2] Corpus Query Processor, cf. Hardie 2012.
[3] Certain texts from a few recent decades in the corpus are still under copyright and cannot yet be made available to everyone for download, but researchers may apply for individual access to the full corpus to be able to query all Present-Day English journal texts from the corpus as well.

Several diachronic and Present-Day English corpora of English contain different types of academic texts, e.g. transcribed lectures and classroom discussions or student essays. The RSC with its published, peer-reviewed journal articles[4] has the advantage of representing a specific text type of professional science communication across a wide time span. There are only very few other English corpora consisting only of published academic texts, e.g. the Corpus of Academic Journal Articles (CAJA; Kosem 2010, ca. 13,000 articles from around 30 disciplines and more than 2000 different contemporary online journals).[5]

The RSC comes with very high usability and accessibility and offers its users various query and visualization options. As a tool available for research, lexicographical and teaching purposes, the RSC provides the possibility to identify and assess the linguistic characteristics of academic journal texts, some of which appear to be highly dependent on the socio-historical context and might have changed considerably over time while others seem to exhibit a certain degree of transhistorical stability. Users of the RSC can explore in many ways how the writing conventions in professional academic discourse have evolved (cf. Menzel 2020). The RSC has been enriched with fine-grained annotations, e.g. for lemmas, parts of speech, text topics obtained by the application of probabilistic topic models (Fankhauser, Knappen and Teich 2016), surprisal as an information-theoretic measure (Degaetano-Ortlieb and Teich 2019), and with metadata that provide users with contextual information (cf. Fischer et al. 2020). Considerable effort has gone into making the RSC a large and high-quality resource with the application of sophisticated OCR (Optical Character Recognition) post-processing and other methods aimed at detecting annotation and tagging errors. The corpus has the advantages of quick data retrieval and powerful search capacities. Two query language options are available: CQP Syntax, a widely used syntax to query corpora, and a simpler version, the Simple Query option. The data may be used, for instance, to compare the use of individual lexical and grammatical items, multi-word expressions, longer phrasal structures and grammatical patterns and discourse structures depending on variables such as author information, text topic or journal type. Concordances and other types of results such as frequency breakdowns, distribution tables and bar charts can be obtained from the RSC via queries from the corpus for particular linguistic

[4] Cf. Fyfe et al. 2019 on the development of peer reviewing practices at the Royal Society journals.

[5] CAJA was compiled for designing a model for a corpus-driven dictionary of general academic English for university students and has restricted access in the commercial corpus management and corpus query software SketchEngine.

constructions. The query results can be sorted and categorized according to the user's own analytic labels and downloaded as a configurable file together with information on metadata and linguistic annotations for further analysis and post-processing. All texts in the corpus are linked to their respective source texts on the Royal Society journal websites so that details about layout elements and figures in the image-based PDF files made from the scans of the original documents and further information on related articles can be retrieved via these links for complementing specific types of analyses. There is no limit to the number of queries for users of the RSC and no need for universities or individual users to pay for access to the available data. Corpus maintenance and the improvement of the quality of the data at all levels is a continuously ongoing process.

Information from the RSC data complements currently available dictionary evidence and lexicographical resources for specialized languages, not only with respect to the historical development of academic vocabulary, but also with regard to modern definitions of terms given in such resources (cf. Menzel 2018). On the basis of the RSC, lists of historical terms involving eponyms with particular structures can be created and sorted according to common linguistic features and / or textual metadata categories (e.g. chronologically or thematically, according to scientific disciplines or text topics). The data enable corpus users to discover diachronic changes of the connotations of eponymous expressions and to reflect on the extent to which their meaning is adequately provided in the definitions and examples in modern general and specialized dictionaries. Relying on existing dictionaries alone without consulting corpus data involves potential difficulties as many dictionaries have their focus mainly on frequent words and selected meanings. Additionally, one may face potentially long searches through entries for the relevant academic sense of a term and may find no relevant usage examples or only rather short ones without any context.

3. The role of eponyms in English scientific discourse

Personal names are an important source of lexeme and term formation elements for vocabulary enrichment in scientific discourse. Various eponyms from academic jargon nowadays may also be found in other text types, in general dictionaries and in common language use (e.g. *Alzheimer's disease*). Some eponyms attested throughout the centuries in the RSC have already been in use in English for a relatively long time, e.g. *Wolffian duct* (also known as *mesonephric duct*), named after a German physiologist from the eighteenth century. This eponym occurs in the RSC from the 1870s onwards to contemporary publications. Other terms have been created more recently from names of early pioneers of modern science, e.g. *Newtonian potential,* used in the RSC from the 1920s onwards (also referred to as *gravitational potential*), or more contemporary scientists (e.g. *Fermi-Dirac statistics*, named after two twentieth-century physicists). Eponymous terms denoting similar concepts can be in competition with one another for a certain time period. In the 1840s, several people of different nationalities claimed to have invented photography. Therefore, various competing terms on this subject coined on the basis of proper names or other word formation elements were in use in the RSC data, e.g. *talbotype* (after William Henry Fox Talbot) and *daguerreotype* (after Louis Daguerre) refer to early photographic techniques.

Many scientific eponyms used in the RSC are derived from names of scholars, inventors and other personalities from the English-speaking world or from European countries. Names from non-European languages as term formation elements in English have become slightly more important in more recent texts in the RSC (e.g. *Killing-Yano tensor* with its second component being the name of a Japanese mathematician, a term used in the RSC from the 1980s onwards, or *Chua's circuit*, named after a Philippines-born Chinese-American, used in the RSC after 1990). English eponymous terms sometimes involve a first name, in the RSC often prefixed with a title, e.g. the *Queen Charlotte form* (a record form in maternity hospitals) or both a first name and a surname as in *Austin Flint murmur* (a heart sound). Some refer to literary, biblical, classical or mythological characters (e.g. *Adam's apple, Achilles tendon, Hipparchian diagram, Lazarus effect, Joseph effect, Pythagorean system, syphilis*). Changing views on how much authority should be given to ancient texts and classical sources is reflected in the development of contexts in which such eponyms are used. Their status as scientific terms and their precise literal or metaphorical meaning for different communities may have changed over time depending on cultural and scientific developments and on processes whereby general or discipline-specific scientific consensus may have been established. There was, for instance, no strict thematic division between historical, philosophical, theological and scientific aspects in texts from scholars such as Reverend Abraham de la Pryme who wrote a letter on fossils as a result of the *Noachian Deluge* that was published in the *Philosophical Transactions* at the beginning of the eighteenth century. Scientists such as Edmond Halley also presented papers to the Royal Society suggesting explanations based on physical and astronomical phenomena for events like the *Noachian Deluge* (also referred to as *Noachical Deluge, Noachian flood, Noah's Deluge,* etc. in the RSC). In later texts from the RSC, Noah is typically referred to in a more metaphorical way in newer terms such as the

Noah Effect in a text on fractal geometry and in *Noachis* in a text on the Martian atmosphere and landscape.

There is a variety of possible semantic relations between scientific eponymous terms and their base words. They all commemorate names; and their main purpose is to be used as scientific terms and to facilitate scientific communication. This does not necessarily mean that their etymological origin is easily recognizable to all language users. Marciano (2009) coined the word ‘anonyponymous’ to refer to the forgotten people behind a number of modern eponyms. Various reasons for thinking critically about eponyms and many stories of scientific renaming have been documented in the scientific and linguistic literature. Different types of people exert the right of coining and evaluating scientific terms and there are various ways in which eponyms spread within the scientific community or to other domains. We may assume that many eponyms in early scientific texts published in English have been formed and introduced by individual eminent scientists or research communities to honor friends, family members, colleagues, pioneering researchers, mentors, sponsors or prominent public figures. One of the main interests of language users who use eponyms or alternative scientific terms is to know how frequently such linguistic items are used and in which contexts and combinations with other structures. Specialized corpus data such as the RSC give us a snapshot of the language from a particular angle. They give us the possibility to zoom in on structures such as eponymous terms and collocations and to look for regularities, but to understand, for instance, why individual eponyms were coined, how they spread within communities or why certain term variants are preferred, we often need to take into account additional sources to complement the information gathered from academic texts in which they are used.

Various terms have found their way into English as internationalisms, e.g. as translations from post-classical Latin terms (*tuba Fallopii—Fallopian tube*). Numerous single and multiword terms were borrowed into English from other, mainly European languages, often in cases where no single descriptive scientific term was yet associated with new inventions or methods. Certain types of historical eponyms can be associated with specific historical strata and reflect cultural and linguistic contacts in specific time periods, e.g. with German, Dutch, Italian or French during the Late Modern English period. For instance, the *Voltaic pile*, an early battery announced by Alessandro Volta in a letter to the Royal Society published in English in the *Philosophical Transactions* in 1800, was referred to with a similar eponym in other languages according to language-specific word formation rules, e.g. French: *pile voltaïque*, German: *Voltasche Säule*, Italian: *Pila di Volta*, Danish: *Voltasøjle*. Some eponyms have become common terms that were or are still widely used while others were mainly used in certain time periods, occurred as occasionalisms or as rather specific terms in very specialized domains (e.g. *Malpighian bodies* occurs particularly often in texts on physiology from the 1840s in the RSC).

With a growing number of eponyms in use and the ever-increasing specialization in the sciences accompanied by ongoing linguistic conventionalization processes in academic English over time, newly introduced eponyms as well as other types of technical terms probably increasingly depend on naming conventions and decisions of communities of English-speaking scientists, nationally organized, specialized assemblies of scientists or international organizations with English as a working language before becoming widely accepted. Nevertheless the exact history of many terms cannot be traced precisely or only through a close examination of historical sources. For instance, at first glance, *Peromyscus stephani (*also known as *San Esteban deer mouse* or *San Esteban Island mouse*) may seem to be named after a person, but it was named after a place, which was itself named after a person. While it is on the one hand often difficult to determine who coined, introduced or popularized an eponymous term first, some eponyms may even raise the question of whether they were meant as a tribute or not. There is, for instance, some controversy about the question whether Linnaeus, in a spirit of ill-will, named species such as foul-smelling weeds or hideous bugs after his rivals to live on with a negative connotation, e.g. the plant *Buffonia,* which is now associated both with name of the taxonomist Georges-Louis Leclerc de Buffon and with ‘bufo’, the Latin word for ‘toad’ (cf. Fée 1858), the beetle *Aphanus rolandri*, whose genus name is taken from a Greek word meaning ‘ignoble’ or ‘obscure’ and whose specific epithet commemorates Linnaeus’ pupil Daniel Rolander, with whom he had personal disagreement, and the bad-smelling weed *Siegesbeckia orientalis*, named after botanist Johann Georg Siegesbeck. More recent examples that have encouraged discussion on whether they were intended as tributes or insults to contemporary individuals are the cockroach *Xestoblatta berenbaumae* named after entomologist May Berenbaum or the slime-mould beetles *Agathidium bushi* and *Agathidium rumsfeldi* named after the American politicians George W. Bush and Donald Rumsfeld. Current eponym creation trends of this type have increasingly raised criticism against giving credit for scientific discoveries to celebrities or politicians, putting scientific names from certain fields on auction or accepting donations in return for name recognition (e.g. for university buildings or professorships) and introducing whimsical and jocular names in biological nomenclature that contain socio-scientific metadata of questionable use and value (Lalchhandama 2014; McClellan

2019). If an official term is named for a person who is still living, it would be difficult nowadays to introduce such a term without their consent and knowledge.

Various recent studies are dedicated to the discussion of arguments for and against using scientific eponyms derived from personal names, particularly in medical fields (e.g. Duque-Parra, Llano-Idárraga and Duque-Parra 2006; Whitworth 2007; Fargen and Hoh 2014; Burdan et al. 2016). There has been some criticism in the literature that eponyms do not always serve as an adequate source for the history of science. Another problem with eponyms is that, in many cases, scientific discoveries are not named after their original discoverer, a phenomenon that has, for instance, been described by sociologist Robert K. Merton and become known as *"Stigler's Law of Eponymy"* (Stigler 1980), which is actually a case of self-eponymization. Naming a scientific discovery after oneself is often considered as a breach of scientific etiquette. Rarely, major scientific breakthroughs are the result of the work of the lone scientist. To acknowledge everyone, for instance, who discovered facets of *Behçet('s) disease*, it has been pointed out that we would have to name it *Hippocrates–Janin–Neumann–Reis–Bluthe–Gilbert–Planner–Remenovsky–Weve–Shigeta–Pils–Grütz–Carol–Ruys–Samek–Fischer–Walter–Roman–Kumer–Adamantiades—Dascalopoulos–Matras–Whitwell–Nishimura–Blobner–Weekers–Reginster–Knapp–Behçet's disease*, but according to Woywodt and Matteson (2007) decisions on which name to choose often depend on having published in a more accessible journal or language, on influence, political considerations and arbitrary factors. There is a long-standing debate among medical scholars and practitioners about the effectiveness of eponyms in comparison to more semantically transparent, scientific descriptions (e.g. *Eustachian tubes* vs. *auditory tubes* or *Koch triangle* vs. *triangle of sinoatrial/atrioventricular node*). The following passage from the *Philosophical Transactions* shows that discussions about using descriptive or eponymous terms were also a subject of debate in other fields represented in the RSC (Ex. 8-1).

> **Ex. 8-1**: "*Dr. Babington has proposed to call this mineral by the name of Wavellite, from the gentleman who discovered it in Devonshire; but if a name founded upon its chemical composition should be preferred, Mr. Davy thinks it may be denominated Hydrargillite."* (RSC, Text ID 109598, Humphry Davy: An Account of Some Analytical Experiments on a Mineral Production from Devonshire, Consisting Principally of Alumine and Water. [Abstract], 1800)

It is not a simple matter to introduce a new name for a well-established term and to phase out the use of an eponym or to create new ones. However, some examples of renaming processes involving an eponym, typically for locations or institutions, are documented in texts from the RSC, cf. Ex. 8-2.

> **Ex. 8-2**: *"When the Von Kármán Gas Dynamics Facility at the Arnold Air Force Station in Tennessee was renamed in his honour on 30 October 1959, he said in his acceptance speech: 'There is no doubt in my mind that this is the greatest honour that I have ever experienced. But I would have thought that the Air Force would have waited a few years until I had the occasion to look back at this from beyond infinity. I think that that is really the tradition, that the scientist or statesman—if something gets dedicated in his name—should already be dead. But I am glad the Air Force managed an exception and gave me the opportunity to accept this honour.* (RSC, Text ID rsbm_1966_0016, Sydney Goldstein: Theodore Von Kármán, 1881-1963 [Obituary], 1966)

Different types of eponyms have raised ethical concerns in recent times, e.g. medical conditions named after specific people, eponyms associated with unethical research practices from the past and historically inaccurate or misleading eponyms. Moreover, several eponyms for different concepts sound very similar (e.g. *Meigs' syndrome, Meige syndrome* and *Meige disease*). Particularly in the medical domain, researchers may have described more than one phenomenon, so that several concepts named after the same person may initially seem semantically related or identical, but actually refer to entirely different conditions (*Pott's puffy tumor, Pott's disease*). Some eponyms may be homonyms and may have different meanings in one or several specialized disciplines. There is also some regional and national variation (e.g. *Sideropenic dysphagia* is both known as *Plummer-Vinson syndrome* named after two Americans and as *Paterson-(Brown-)Kelly syndrome* named after British doctors), which may lead to confusion or make it necessary to mention several alternative terms in academic texts for clarification. There are eponyms with names that are not consistently used in the same order (e.g. *Creutzfeldt-Jakob disease* / *Jakob-Creutzfeldt disease*), while other eponyms may have possessive and non-possessive form (e.g. *Parkinson's / Parkinson disease*) or different spellings for personal names (e.g. *Bechterew's / Bekhterew's disease*[6]). Additionally, the association between a person and a concept may get lost over time so that the informational content of eponymous terms can become low for those who use them without being aware of their etymological and historical

[6] Over the years and depending on regional preferences, many other eponyms and descriptive terms have also been used to refer to this disease, cf. Khan and van der Linden 2019.

roots. Nevertheless, eponyms can serve as convenient shortcuts for concepts whose exact characteristics may be difficult to remember or refer to with a reasonably short descriptive equivalent. Moreover, various arguments against potentially misleading eponyms such as the occasional existence of terminological, regional and orthographic variation and disagreement on the most appropriate term from a historical, ethical and scientific point of view may also be brought forward against the results of other word formation processes. Long descriptive terms may initially be semantically more transparent than eponyms, but they have a strong tendency to become known primarily by their abbreviations or acronyms, which again may lead to misunderstandings. In spite of the long list of arguments that have been raised against eponyms and recommendations by organizations such as the World Health Organization to use descriptive names for new concepts such as new infectious diseases, scientific eponyms continue to be widely used in the English language and in other languages in scientific fields in which their use is a time-honoured tradition. There are similar discussions from a scientific and linguistic point of view on the implications of using the names of geographical places and social, national, regional, professional or ethnic groups etc. in scientific language. Such forms do not seem to be marching towards extinction either although individual forms lead to criticism from various sides as well. The editor of a scientific journal, the *Lancet*, for instance, had to apologize for publishing a 2010 paper in which microbiologists used the term *New Delhi Metallo-Beta Lactamase-1* for an enzyme.

4. Existing lexicographical descriptions of eponyms

The high number of studies on eponyms may leave the impression that there is an abundance of literature on this topic. This impression is reinforced by the fact that the term ‘eponym’ has been used in literature with several, distinct meanings. It may refer to a lexical item derived from a personal name, to the personal name that serves as the base or to a person from whose name a lexical item is derived. Some would include lexemes derived from place names within the group of eponyms or proper nouns that have become common noun, e.g. brand names (‘proprietary eponyms’). Therefore, there are also several possible meanings of the adjective ‘eponymous’ when combined with nouns such as *word, term, expression, phrase, dictionary, person, hero, scholar, disease, brand* etc. One could even say that scientific areas such as medicine or psychology are very eponymous subjects. The decision on where to draw the exact line for including forms of a certain type into the definition of eponyms or into a dictionary remains a matter of preference and purpose; and no single theoretical or definitional approach will serve all purposes.

Several academic fields are treated again and again in the literature on eponyms, e.g. anatomy, medicine, psychology and biological nomenclature, while other productive domains for English eponyms are not covered to the same extent. Various commonly and frequently used eponyms may be found in general English dictionaries. As the proportion of technical and scientific terms in general dictionaries has risen in recent times, some specific eponymous terms from scientific discourse now also have entries in resources such as the OED3 (e.g. *Wegener's granulomatosis*).[7] There are some English eponym dictionaries which are not restricted to any particular discipline or field and which are dedicated to eponyms used in Present-Day English. Freeman's (1997), Elsevier's (Letusé La O 2001) and the Chambers Dictionary of Eponyms (Manser 2004), for instance, provide alphabetical lists for eponyms along with some etymological aspects, sample sentences and much biographical, historical and sometimes popular-scientific information on eminent personalities and terms named after them. The numbers of their entries range from several hundred up to almost three thousand. Contemporary English eponyms as a subset of scientific vocabulary collected by lexicographers and scientists are covered in a number of monolingual specialized resources which differ considerably with regard to their presentation (specialized dictionaries and glossaries of scientific eponyms or of different types of scientific terms including eponyms, encyclopaedias with eponyms from certain fields presented in the context of biographical sketches and historical information on scientific breakthroughs and naming processes). Eponymous terms from medical terminology and biological nomenclature are relatively well documented in lexicographical resources and have received considerable interest from the research community. There are eponym dictionaries dedicated specifically to the names of birds, mammals, reptiles, amphibians or plants (Beolens and Watkins 2003; Beolens, Watkins, and Grayson 2009, 2011, 2013; Quattrocchi 1999), dictionaries and encyclopaedias of eponyms used in medical discourse or in specific subdisciplines from the medical sciences (e.g. Dobson 1946; Bartolucci and Forbis 2005; Baskett 2019). Some publications deal with scientific terms named after people depending on specific social

[7] This term for a disease can be found in the currently valid international classification scheme ICD-10 maintained by the World Health Organization as well (https://www.icd10data.com), but its use is now discouraged by some medical societies that recommend to phase it out (due to Friedrich Wegener's controversial past as a pathologist during the 1930s) and to use the descriptive term *granulomatosis with polyangiitis* instead (cf. Falk et al. 2011).

or cultural aspects, e.g. minerals named after Swiss individuals (Roth 2006), eponymous terms in dermatology or ophthalmology named after women (Bader and Shipman 2015; Van Tassel et al. 2018) or medical eponyms derived from names of literary characters (Rodin and Key 1989). Popescu (2009a) lists around fifty lexicographic products for English eponyms, most of which either cover eponyms in general English or medical terms.[8] The need for investigating English eponyms in various scientific areas that are comparably under-researched from a linguistic point of view and less well documented with respect to eponymous units functioning as terms of science than some types of life sciences becomes obvious in studies by Popescu and Maftei (2003, 2005) and Popescu (2007, 2009b). It is difficult to find studies and mono- or bilingual resources on specialized eponyms in various fields such as chemistry, mathematics, physics and engineering.

In translation studies, the challenges of dealing with the translation of eponyms have rarely been examined in detail. Newmark (1988, 146-147) briefly mentions eponyms (in his definition all derivations from proper names, including toponyms and brand names) as a subtype of neologisms that may involve difficulties particular in the translation of media texts. He presents one example of a technical term from athletics (*Fosbury flop*) and points out in a rather cursory manner that in cases where an eponym does not refer directly to the person the translator may have to make the target text more explicit, which also depends on whether the text is targeted at specialists or non-specialists. One might assume that scientific eponyms often function as internationalisms with direct equivalents in other languages, at least in most European languages (cf., for instance, the historical example of the *Voltaic pile* given above). However, the above-mentioned studies by Popescu and Maftei on contemporary texts from highly specialized registers revealed various translation difficulties for scientific eponyms due to the lack of suitable resources and, in many cases, of direct equivalents for language-specific eponyms in this language pair. In such situations, translators may find it difficult to grasp the exact meaning of specialized eponymous terms, to identify single- or multi-word eponymous or non-eponymous translation equivalents and to recognize cases where it is necessary to resort to specific explicitation and reformulation strategies.

Corpus studies on eponyms offering routes into lexicographical resources are rare. It is hard to find specialized, searchable data that may serve to underpin assumptions and intuitions on frequencies and the

[8] Popescu, who has published on a number of eponym-related topics, is currently contributing to the compiling of an English-Romanian dictionary of eponyms (cf. https://www.cambridgescholars.com/t/AdvisoryBoardFloriana, accessed September 06, 2020).

productivity of eponymous terms and their underlying word formation rules. Phrasal and collocational combinations as well as possible ambiguities of eponyms are not covered sufficiently in existing resources and might be something dictionary users would like to see included in a specialized dictionary. One of the few corpus-based studies on eponyms that have been conducted is by Cabanac (2014) who described semi-automatic extraction strategies of a specific type of English eponyms in a corpus of around 800 unannotated full-text articles published in the journal *Scientometrics* between 2010 and 2013.

The results of various analyses and the discussion of examples from the RSC in the next section demonstrate the value of the RSC as a corpus resource for the identification and collection of eponyms. Academic and scientific terms involving eponyms can be retrieved according to their use in specific time periods or in relation to specific topics and scientific disciplines. The results may be used to complement synchronic and diachronic corpus-based dictionaries and resources for specialized languages to be used, for instance, in teaching, studying and research with regard to general or discipline-specific English for Academic Purposes and for Specific Purposes. They may also serve as a basis for reconstructing cross-cultural aspects of scientific eponyms.

5. Identifying eponyms and typical word-formation patterns in the RSC

Eponymous terms in general do not form a structurally homogeneous group in English and have different morphosyntactic features. Word formation processes for eponyms in the RSC involve a wide range of structures ranging from single-word units of different types such as conversions, derivations or abbreviated forms to multiword expressions and longer phrasal and collocational patterns. In this section, various examples will be given to illustrate these different linguistic patterns. A particular focus will be placed on the results of the analyses of the most typical, regular term formation patterns for scientific eponyms in this dataset.

For many single-word eponyms occurring in the RSC no systematic and dominant word formation pattern can be attested. Some single-word eponyms are part of larger groups of lexemes used across the dataset which are derived from the same proper name, although not all of them are closely semantically related to each other (e.g. *galvanization, galvanoscope, galvanoscopic, galvanometer, galvanic, galvanism, galvanoplastically*). As the RSC journals cover a broad range of topics, single-word eponyms occurring in the RSC are quite diverse and often have no specific morphemes or other structural elements in common apart from their reference to the name of a particular person. They were formed from proper names or alternate forms or derivatives of proper names recategorized into common nouns. Examples of one-morphemic proper noun to common noun conversions, sometimes referred to as commonization or antonomasia, are units of measurement (*ohm, volt,* etc.*)*. They can also be the result of shortening processes applied to longer patterns (e.g. *Diesel engine*, used from the 1920s onwards in the RSC, after the 1940s increasingly without capital letter for the proper name and after the 1970s frequently without a subsequent noun). Other types of eponymous lexemes are multimorphemic nouns with various combining forms and/or affixes (*galvanization, roentgenology, cobezoutiants, Hobbist, Socinians)*, but also sometimes verbs or adjectives (*pasteurize*, *galvanoscopic*). A few types of scientific single-word eponyms follow regular patterns, e.g. the addition of a certain suffix to a personal name, in analogy to similar terms named after localities, physical properties, etc. Nomenclatural botanical and zoological eponyms named after people typically end in Latin genitives (*Chlamydomonas reinhardtii*), minerals named after persons end in *-ite (ankerite)* and chemical elements named after people end in *-(i)um* (*curium*). Due to their diversity, single-word eponyms in the data may primarily be identified with queries for specific lexemes, e.g. those derived from specific names of individuals. Looking for terms derived from the name of Isaac Newton, for instance, will lead to both single- and multiword units such as *Newtonianism* (used from the 1890s onwards) or *Newtonian theory* (used from the 1740s onwards) and complex phrases such as the *quasi-Newton-Raphson conjugate gradient minimizer VA09A* (in a text from the 1990s). Other examples include the *Newton telescope*, the unit *Newton* and its abbreviation *N, Isaac Newton Institute for Mathematical Science*, *Newton Raphson method*, *non-Newtonian fluids*, *hybrid Newton/time-integration procedure* and *pseudo-Gauss-Newton procedure.* In the case of these examples, all eponymous terms identified in the RSC in early times and in more recent texts have been derived from the name of Isaac Newton. As various scientists and personalities mentioned in the corpus may have the same surname (Isaac Newton, Alfred Newton, William E. Newton, etc.) it might be necessary to check whether the eponyms in connection with selected names all refer to the same personality. If a name is less well known nowadays or if different people share the same name, this might give rise to folk etymologies and it might be difficult to establish a correct etymological connection.

The more complex forms of terminological eponyms and those with prefixes are typically found among the more recently created terms. The usage of abbreviations and initialisms, blends, clippings, conversions, backformations and backronyms from eponyms also reflects rather recent word formation trends. In various contemporary academic texts from other resources short forms such as *h*[irsch]-*index*, *Alzheimer‘s* [disease], *C*[aesarean]-*section*, *APGAR* [test/score][9] are popular forms but in the RSC data such shortened forms on the basis of eponyms seem to occur only occasionally (e.g. the *H.K.R.R. formula*, cf. Ex. 8-3 and the *KAM theorem*, discussed in Section 6 below) and they are difficult to identify systematically via corpus queries without checking the full wording of such forms elsewhere. Many academic and scientific terms in the RSC honoring the name of people are multiword eponyms following regular term formation pattern. The most important patterns of eponyms for this dataset are terms containing a suffixed proper adjective ending in *-ian, -ean, -an, -ic,* or *-esque* followed by a noun (e.g. *Brownian movement, Napierian logarithm, Voltaic sparks*), proper noun(s) + noun(s) structures (e.g. *Basset force, Newton cooling coefficient, Mann-Whitney U test, Einstein-Yang-Mills-Higgs model)* and noun phrases containing a possessive marker (e.g. *Hadley's quadrant, Steiner's curve, Fahrenheit's scale*). Various rather general nouns occur regularly in eponyms (e.g. *model, law, method, theory*). Such sets of eponyms with similar structures can be systematically retrieved with appropriate CQP-queries.

Additionally, an interactive visualization tool for word embeddings provides useful co-occurrence information and may help to find groups of particular forms of proper names and other potentially relevant linguistic items that are semantically similar such as groups of abstract or concrete nouns that occur regularly in eponyms. Fig. 8-1 shows an example of the semantic neighbourhood of *'Newtonian'* in the 1780s (on the right) and its use over time in frequency per million tokens (FpM; on the left). In the visualization it co-occurs with various abstract and concrete nouns and with other proper adjectives used in similar eponymic word formation patterns in the corpus data from this time period (e.g. *Newtonian telescope, Newtonian reflector, Gregorian telescope*). The bubble chart represents the semantic space of words, with the size of bubbles proportional to the square root of the relative frequency in the chosen interval. Frequency change is visualized by different colors. Fig. 8-1 appears as a grayscale image in this volume, therefore negative values are marked with a dashed circle.[10]

[9] Named after Virginia Apgar, but now mainly known as mnemonic backronym for *Appearance, Pulse, Grimace, Activity, Respiration* as a measure of a baby's condition after birth.

[10] For the original see: http://corpora.ids-mannheim.de/diaviz/royalsociety.html

Paradigmatic change can be assumed for islands with words of similar colors. (cf. Fankhauser and Kupietz, n.d.).

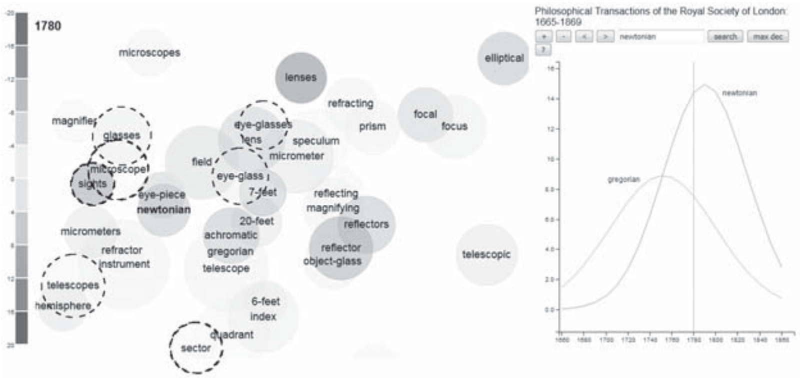


Fig. 8-1: Zoomed-in region of semantic neighbourhood of *'Newtonian'* in the 1780s.

In this context, it is worthwhile to note some challenges of formulating optimized queries for the automatic retrieval of eponyms. One difficulty arises from the fact that the same types of corpus annotation methods were applied to different phases of English spanning over 300 years. Therefore, various query refinements are necessary depending on characteristic linguistic features of specific time periods. For instance, capitalization and proper noun part-of-speech tags can often be good indicators for potential eponyms, but particularly in the data from before 1750, many ordinary nouns that authors or printers wanted to make more prominent or that belong to certain semantic categories were also written with initial capital letters and may occasionally and misleadingly have been assigned a proper noun tag. Variation in hyphenation, spelling, foreign elements, etc., also needs to be taken into account, e.g. to find variants such as *Leibnitian Doctrine / Leibnitzian doctrine, Schottky-barrier height / Schottky barrier height, Max-Planck-Gesellschaft / Max Planck Society*). Some derivations have slightly different spellings from the personal name they are related to (e.g. *Horroxian,* derived from Horrocks, *Cartesian* derived from Descartes). Other similarly suffixed, capitalized adjectives that are irrelevant for the search for eponymous scientific terms have to be excluded in refined queries, e.g. by excluding words such as *Australian* or *Precambrian*.

A higher precision ratio, at the cost of lower recall though, can be achieved by formulating even more restricted searches, e.g. for proper nouns preceded by a definite or indefinite article and / or followed by a particular noun that occurs in many eponyms such as *law, virus.* This method also helps to identify surnames that have homonyms which are not personal names, e.g. *Wien law, Friend virus.* Nevertheless, personal names remain difficult to differentiate from homonyms automatically without manual verification (e.g. the *Beer law*, which has nothing to do with the German beer purity law in the RSC data, but with a mathematical equation named after August Beer), particularly if several words in a term or an entire phrases are capitalized (e.g. *Green Theorem*, *BORN APPROXIMATION*).

Non-refined broader queries will lead to a higher recall, but also to many irrelevant examples that subsequently have to be sorted out from the results. Some manual filtering is necessary with regard to individual words with structurally similar features to personal names in eponyms that are difficult to anticipate via a predefined list due to their variety (e.g. proper nouns denoting localities as in *Baginton-Lillington gravels* or *Greenwich table*) and of structurally identical, ambiguous forms. The adjective *Hessian,* for instance, is potentially related to terms derived from the mathematician Otto Hesse (*Hessian matrix*) or from the German region of Hesse (*Hessian crucible*). *Gregorian* may refer to the astronomer James Gregory in terms such as *Gregorian telescope* or to Pope Gregory when talking about the *Gregorian calendar*. Some surnames may look like first names (e.g. *Martin Ice Rise, Henry Mountains*). Some structures that could be potential candidates for eponyms occur in noun phrases which do not function as a technical term (e.g. *the Aristotelian definition of the soul, Helmholtz's theory of the objective existence of partial tones, Wullner's experiments on the elasticity of steam*). Noun phrases with possessives that do not function as terms can in many cases be excluded via targeted queries as they are often followed by a prepositional phrase, typically an *of*-phrase. More specific queries were developed in the course of this study to enhance precision and recall for the search results in the RSC.

6. Analysis and results

The results that will be presented in this section have a focus on the most popular term formation patterns in the RSC. Based on the extracted results of eponyms with different morphosyntactic features it can be said that early formed eponymous terms in the RSC are predominantly adjective + noun patterns which were most productively coined in the first 200 years of the data. In certain periods such as between 1850 and 1900 as well as during the last 50 years of the corpus data, a particularly sharp increase with regard to the frequency and productivity of eponyms used in academic and technical vocabulary can be observed. Proper noun + noun patterns were attested as the most productive word-formation type in data from the last 130 years.

Fig. 8-2 shows the usage of adjective + noun eponyms across the most important disciplines represented in the corpus data throughout the first 200 years (1665-1869). In their early stages, the journals whose texts are part of the RSC used to cover essentially all branches of science known at this time. Therefore, academic disciplines for this time span have been approximated by grouping journal texts together according to their topic labels in the textual metadata that were obtained by the application of probabilistic topic models. Optics, mechanics, electromagnetism and thermodynamics, for instance, as a group of topics are treated as a subcorpus of physics research articles here.

Fig. 8-2: Adjective + noun eponyms in first 200 years of RSC across disciplines in FpM and normalized in relation to the respective size of each domain-specific subcorpus.

Relatively similar trends can be observed in all disciplines, but biology shows a prominent peak in the eighteenth century. This can be explained by the fact that, especially during the eighteenth century, individual biology texts in the data are characterized by a high token frequency of adjective + noun eponyms. The terms show specific distribution patterns within and across documents. It is possible to use a specific distribution breakdown in CQPweb to display two lists of text-frequency extremes showing the texts with the most frequent occurrences for a pattern or a term and those texts in which such a pattern or term is found least frequently (but at least once). Outliers, i.e. texts with extreme values, can subsequently be singled out easily if one wants to exclude them from the analysis. The type frequencies for this term formation pattern are actually lower in biology during this time span than in other disciplines. Due to the broad range of subjects covered in these early journal texts, many academic eponyms occur only occasionally or as hapax legomena. In total, around 3000 different adjective + noun types derived from ca. 160 personal names were identified in the first 200 years of the dataset. Around 80 different eponymous adjectives are used more than three times in adjective + noun constructions during this time span. After the journals had split into more specialized series covering biological sciences and mathematics and physics separately, not many new terms with this word formation pattern were introduced (e.g. *non-Newtonian fluids* is used only after 1940), but many of the terms that were coined early continue to exist also in the newer data (e.g. *Torricellian vacuum* occurs in RSC texts from the 1750s until the 1980s). The word cloud in Fig. 8-3 captures the most frequent adjectives used in eponyms that were already in use in Early and Late Modern English texts in the RSC.

Fig. 8-3: Most frequent eponymous adjectives in adjective + noun patterns in RSC texts from Early and Late Modern English.

It may not come as a complete surprise that many proper adjectives in Fig. 8-3 are derivations from British and European names and that *Newtonian, Bakerian* and *Croonian* are among the most frequent ones in this dataset. However, this word cloud may serve as a simple example of how information on eponyms may be presented visually for lexicographical and teaching purposes. It may illustrate and summarize some aspects and relations better than a table of data and could easily be integrated into an interactive application to investigate the data in more detail and context. If we shift our focus from the most frequent eponymous adjective in Fig. 8-3

(*Newtonian*) to the whole adjective + noun structure it is interesting to see, for instance, that this adjective does not occur within the top 15 eponymous adjective + noun collocations of the full RSC and the first 200 years of the data (cf. Table 8-2). Zooming in on the last 50 years of the data in Table 8-2, it occurs there only in two collocations, ranking 11th and 15th. The adjectives we see in Fig. 8-3 are combined with a great variety of specific and abstract nouns in the data (e.g. *Newtonian fluid / theory / telescope / reflector*) or they occur only with one or very few particular nouns in specific terms that form frequently used collocations (e.g. *Haversian canals / systems*).

	Top 15 adj. + noun eponyms in full RSC V5.1.0 (size: 297,965,584 tokens)	**FpM**	**Top 15 adj.+noun eponyms in last 50 years 1950-1996 (size: 173,147,836 tokens)**	**FpM**	**Top 15 adj.+noun eponyms in first 200 years (1665–1869) (size: 31,952,725 tokens)**	**FpM**
1	*Bakerian Lecture*	3.9	*Croonian Lecture*	1.65	*Fallopian Tube*	8.32
2	*Croonian Lecture*	3.7	*Bakerian Lecture*	1.41	*Bakerian Lecture*	5.41
3	*Fallopian Tube*	1.9	*Brownian motion*	1.10	*Graafian vesicle*	4.07
4	*Brownian motion*	1.8	*Gaussian distribution*	1.09	*Croonian Lecture*	2.72
5	*Gaussian distribution*	1.5	*Smithsonian Institution*	0.71	*Malpighian bodies*	2.53
6	*Smithsonian Institution*	1.1	*Cartesian co-ordinates*	0.66	*Savilian Professor*	2.47
7	*Brownian movement*	1.1	*Malpighian tubules*	0.64	*Torricellian vacuum*	2.35
8	*Malpighian tubules*	0.9	*Gaussian curvature*	0.53	*Fallopian Tubes*	2.25
9	*Fallopian Tubes*	0.9	*Bodleian Library*	0.53	*Eustachian Tube*	2.25
10	*Gaussian curvature*	0.8	*Maxwellian distribution*	0.50	*Haversian canals*	2.22
11	*Maxwellian distribution*	0.8	*Newtonian fluid*	0.42	*Napierian Logarithm*	1.44
12	*Weberian ossicles*	0.8	*Hamiltonian systems*	0.40	*Haversian systems*	1.34

13	*Savilian Professor*	0.7	*Cartesian components*	0.38	*Fullerian Prof.*	1.12
14	*Eustachian Tube*	0.7	*Lagrangian density*	0.38	*Eustachian tubes*	1.03
15	*Bodleian Library*	0.6	*Newtonian theory*	0.33	*Wolffian bodies*	0.97

Table 8-2: Top 15 eponymous adjective + noun patterns in the RSC and in selected time periods.[11]

Table 8-2 shows that the most frequently used adjectival forms in the top 15 eponymous adjective + noun patterns end in the suffix *-ian* and not in other forms such as *-ean, -an, -ic* or *-esque.* As could have been expected they often refer to influential historical figures of academic life and / or the Royal Society itself and to pioneers of modern science. Some perpetuate throughout the years as proper names for academic institutions or as terms for particular events, honors or sponsorships reflecting historical and contemporary aspects of academic life in Britain or abroad (e.g. *Bakerian Lecture*, instituted in consequence of a bequest of Henry Baker to the Royal Society or *Jacksonian prize, Fullerian Professor*, *Linnean Society* and *Smithsonian Institution*).

As can also be seen in Table 8-2 the normalized frequencies of the most important adjective + noun eponyms during the first 200 years are higher (up to 8.32 per million tokens) than the average normalized frequencies of the most frequently used forms in the whole dataset (up to 3.9 in FpM) and much higher than those from the last 50 years of the data (up to 1.65 in FpM). Many frequently used terms following this pattern that were already attested during the first 200 years also belong to the most prominent adjective + noun eponyms in the RSC in general and in its most recent texts. For some concepts with adjectives, there are differently structured term variants in the data, e.g. *Newtonian / Newton's / Newton rings, Bodleian / Bodley's / Bodley Library*). Often formally similar equivalents in other European languages exist for English terms with adjective + noun structures and various translation equivalents also involve adjectival forms (e.g. the *Voltaic pile* mentioned above in Section 3, the *Lucasian Professor* at the University of Cambridge [French: *professeur lucasien*, Italian: *professore lucasiano*, German: *Lucasischer Professor*]

[11] Both capitalized and uncapitalized forms as spelling variants have been included in these results. Other typographic variants have not been considered here as they occur much less frequently (e.g. *Falloppian*, after Falloppio (Fallopius), used as an early spelling variant of *Fallopian*).

and the *Bodleian Library* [French: *Bibliothèque Bodléienne*, post-classical Latin / Italian / Spanish: *Biblioteca Bodleiana*]). Eponyms with similar or different structures may serve as terms from the same field that are semantically related through lexical relations such as synonymy, antonymy or hyponymy. Eponyms therefore co-occur in various texts, e.g. in the phrases *"Pythagorean and Platonic mathematics", "a Confirmation of the Copernican System against the Ptolemaic and Tychonic"* or *"Schottky barriers and Ohmic contacts"* or in the examples given in Fig. 8-4 from texts from the physical, mathematical and engineering sciences in the *Philosophical Transactions* A (rsta) and *Proceedings A* (rspa), in which *"Einsteinian"* followed by a noun co-occurs with various other terms involving proper adjectives (*Newtonian, Riemannian*).

rspa_1966_0116	avoid such elaboration of expression , let us decide to regard the	**Einsteinian metric**	as the basis for our geometrical terms , and think
rsta_1977_0035	this level uncover a change as radical as that from Newtonian to	**Einsteinian theory**	? 6 . The gravitational parameters The now classi
rspa_1978_0146	for supposing that (2) and (3) contain both	**Einsteinian gravity**	and other fields , in particular Yang-Mills gauge f
rspa_1978_0146	a super-Rieman- nian manifold to obtain less virulent . behaviour than strict	**Einsteinian gravity**	. The alternative approach of super-gravity (Wes
rspa_1992_0111	Damour , T. 1987 The problem of motion in Newtonian and	**Einsteinian gravity**	. In Three hundred years of gravitation (ed. S. W

Fig. 8-4: Examples from a concordance search for terms with the pattern *"Einsteinian"* + noun.

Surprisal, an information-theoretic operationalization, has been calculated for various units and annotated in the corpus to measure the probability of a linguistic unit to occur in a given textual context. Surprisal values for proper adjectives in the data are typically very high.[12] The surprisal values of the noun following the adjective show higher variability depending on the degree of term conventionalization. The adjective *galvanoscopic*, for instance—consisting of two combining forms, one of which has been derived from a proper name, and a suffix—most typically occurs before the word *frog* that comes with a relatively low surprisal value after this adjective. Sometimes it is also used before words such as *effects, nerve, limb* or *leg*. The most typical contextual pattern for this adjective, taking into account its local context, is "*nerve of the galvanoscopic frog"*, in which this adjective has a lower surprisal value than in other contexts. From an information-theoretic point of view, eponyms tend to be

[12] Their Average surprisal (AvS) value is usually higher than 14 in the data. AvS is a measure of the amount of information transmitted by a linguistic unit (e.g. word), averaged over all its instances in a given time period in relation to its textual context. Here the information captured by context refers to a local context of the three preceding words (cf. Kermes and Teich 2017).

semantically effective, informationally dense expressions. They seem to be used as a signal to guide attention, particularly those derived from rarer names. Surprisal values can be downloaded for selected query results or visualized with a tool implemented in CQPweb (Fischer, Fankhauser and Teich 2017). Fig. 8-5 shows an example from a visualization of surprisal values for an extract from a corpus text from the 1840s.

Fig. 8-5: Visualization of higher and lower surprisal values by the use of larger and smaller font sizes in a passage from the RSC.

One reason for the decreasing importance of the adjective + noun pattern over time is that suffixed adjectives cannot easily be combined with all types of names (e.g. names from non-European languages), with terms referring to more than one person or with other multi-word eponyms and other recent word formation patterns which are increasingly found in the more contemporary data. Possessive forms with the pattern proper noun + apostrophe + s (e.g., *Fehling's solution, Fechner's Law*) are used occasionally throughout the entire dataset, but did not become very productive over time. This can probably be explained by their strong indication of ownership and their potential syntactic ambiguity that makes them not easily distinguishable from true possessives. They most typically occur as less frequently used terms variants (e.g. *Down‘s syndrome* vs. *Down syndrome* or the above-cited examples of *Newton's rings* and *Bodley's Library*), preferably with mono- and bisyllabic names *(Chua's circuit, Ohm's law*) and with a rather limited number of abstract nouns that are frequent in eponymous concepts (e.g. *law*). The possessive form may also serve as a grammatical equivalent of formerly used Latin terms (*Herophilus' press / Torcular Herophili*). Only a few eponyms in the RSC come in the form of periphrastic structures, e.g. *of*-genitives such as *Circle of Willis*. Such forms are also sometimes used as contemporary or historical term variants, e.g. *law of Ohm* as an alternative for *Ohm‘s law*. Particularly in the medical domain inconsistencies due to term variants with possessive forms are increasingly frowned upon.

Patterns consisting of one or two proper noun and a common noun become established as the most popular term formation patterns for eponyms in the RSC. Earlier terms of this pattern in the data are often formed on the basis of the name of only one person that may occur in in various different terms (e.g. *Coulomb potential / law / force,* etc.). Hyphenated double eponyms (e.g. *Navier-Stokes equations*) become a very typical term pattern over time. Sometimes these double eponyms indicate a type of collaboration between the people after whom a term is named, but many types of relations between the proper noun components are possible. The *Lotka-Volterra model*, for instance, has presumably been developed independently by Alfred Lotka and Vito Volterra. Other terms consisting of two names indicate that an original scientific concept, model or invention was modified by another scientist whose name has been added later to the term or that several concepts have been brought together in one term. Occasionally both nouns refer to the name of one person (e.g., *Bence-Jones proteins*, named after Henry Bence Jones). Almost all eponyms attested in the RSC are named after men, surprisingly also in the most recent data. The rare cases of eponymous terms named after women can typically be found in the life sciences (e.g. *Barbara M. Gilchrist haemoglobin value*, used in an RSC text from the 1950s or *Friend virus*, named after Charlotte Friend, used in the RSC from the 1970s onwards) or they refer to exceptional, powerful women who played an important role in society (e.g. *Queen Victoria Scholarship*).

Looking at the number of existing publications and resources on eponyms in medical terminology, anatomical and biological nomenclature, one might think that the *Philosophical Transactions* and *Proceedings B* on biological sciences in the RSC (texts from the late 1880s to 1996, ca. 100 M tokens) would be characterized by a higher frequency of eponyms than the A series of the *Philosophical Transactions* and the *Proceedings A* (texts from the same time period, ca. 120 M tokens). Interestingly this is not the case as the journals covering mathematics and physical sciences have turned out to be more 'eponymophile' with regard to the frequently used proper noun + noun pattern. There were, for instance, 9,324 different types of hyphenated double eponyms with 25,958 occurrences in total in the A series journals. Normalized in relation to the size of this corpus part this corresponds to 212 occurrences per million tokens. In the B series journals, 3860 different types could be identified, but here the type / token ratio was higher. There were 7654 occurrences in total, which corresponds to a FpM of 76 in relation to the size of the corpus part with these journals.

	Hyphenated double eponyms in RSC (Philosophical Transactions and Proceedings A)	**abs. frequency**	**Hyphenated double eponyms in RSC (Phil. Transactions and Proceedings B)**	**abs. frequency**
1	*Navier-Stokes equations*	731	*Creutzfeldt-Jakob disease*	148
2	*Fabry-Perot interferometer*	183	*Epstein-Barr virus*	78
3	*Hartree-Fock calculations*	177	*Rohon-Beard cells*	73
4	*Fermi-Dirac statistics*	156	*Tamm-Horsfall glycoprotein*	59
5	*Wiener-Hopf technique*	141	*Kolmer-Agduhr cells*	56
6	*Einstein-Maxwell equations*	133	*Bence-Jones proteins*	51
7	*Franck-Condon factors*	133	*Hodgkin-Huxley equations*	49
8	*Orr-Sommerfeld equation*	128	*Rohon-Beard neurons*	41
9	*Jahn-Teller effect*	126	*Levin-Wyman ergometer*	37
10	*Navier-Stokes equation*	126	*Ringer-Locke solution*	35
11	*Euler-Lagrange equations*	117	*Krebs-Ringer solution*	33
12	*Fokker-Planck equation*	114	*Mann-Whitney U-test*	33
13	*Hartree-Fock approximation*	109	*Michaelis-Menten kinetics*	32
14	*Tollmien-Schlichting waves*	107	*Hardy-Weinberg equilibrium*	30
15	*Klein-Gordon equation*	106	*Mann-Whitney test*	30
16	*Born-Oppenheimer approximation*	103	*Sprague-Dawley rats*	29
17	*Killing-Yano tensor*	102	*Lotka-Volterra model*	28
18	*Franck-Condon principle*	98	*Brown-Sequard Phenomenon*	27
19	*Hartree-Fock equations*	97	*Clare-Bishop area*	27
20	*Runge-Kutta method*	97	*Gerstmann-Straussler syndrome*	27

Table 8-3: Top 20 eponyms in RSC consisting of two hyphenated proper names followed by one noun.

Table 8-3 gives an overview of the top 20 forms of this pattern. Different forms of the same term (e.g. spelling variants or singular and plural forms) are treated separately in these lists, but may also easily be grouped together.

The 50 most frequently occurring forms of this pattern in the A journals predominantly include rather general nouns related to abstract mathematical concepts, e.g. sixteen terms denote equations, many others contain nouns such as *approximation, calculations, distortion, distribution, effect, factor, instability, law, method, statistics* or *theory*. The fifty most frequent terms in the B journals have lower absolute frequencies than those from the A journals and include a greater diversity of different abstract and concrete, sometimes rather specific nouns, but also many nouns related to anatomy and diseases (*area, cells, disease, equation, glucose, glycoprotein, method, phenomenon, solution, syndrome, test, virus,* etc.).

Recent scientific advances in many fields are often associated with groups of researchers. This is well reflected in the increasing number of multi-author papers in the RSC and in the creation of more complex structures in the more recent corpus parts, e.g. triple eponyms or even longer patterns such as *Fresnel-Kirchhoff-diffraction theory* or *Rice-Longuet-Higgins-Nayak analysis.* They may become popular in the form of initialisms, e.g. *K.A.M. / KAM theorem* or simply *K.A.M.* is used in the data, but not the term Kolmogorov-Arnold-Moser theorem). Only one instance of a hyphenated four names + noun pattern can be attested in the B journals and it occurs only in the most recent decade of the corpus (*Hines-Jonker-Taylor-Zeeman dynamics*), while there are several such forms in the A journals (*Heither-London-Slater-Pauling method, Einstein-Yang-Mills-Higgs model, Einstein-Yang-Mills-Higgs system, Lande-Fermi-Goudsmit-Segre formula, Einstein-Yang-Mills-Higgs equations, Mansfield-Rhim-Elleman-Vaughan 8-pulse, Pauli-Heisenberg-Russell-Hund theory,* etc.). The longer and the more complex a word formation pattern for eponyms becomes, the higher is its ratio of different constructions or types to total occurrences. Nevertheless, eponyms combining more than two names typically contain a very general common noun. Hyphenated triple eponyms have around 400 different types with ca. 600 total occurrences in the mathematical and physics journals of the A series and around 200 types with ca. 300 occurrences in the life sciences journals of the B series. Almost no patterns with four or more names are used more than once. Constructions with five to six personal names only occur in the A journals and are used from the 1970s onwards (*Caudrey-Dodd-Gibbon-Sawada-Kotera equation, Hertz-Bromwich-Debye-Whittaker-Penrose potentials, Smoluchowski-Cunningham-Knudsen-Weber-Millikan expression, Einstein-Stokes-*

Cunningham-Knudsen-Weber-Millikan relations). There is generally a high number of eponyms occurring as hapaxes in the later texts, which may be interpreted as an indicator of increasing productivity of this word formation process (cf. Baayen and Lieber 1991), but it also has to be borne in mind that the amount of texts is particularly large for the last 100 years of the data and one might want to use a different, more balanced corpus version of the RSC with equal amounts of texts per time period in a more detailed analysis of the productivity of word formation patterns.

The results from the extraction and corpus-based analysis of eponyms for various patterns can complement historical and contemporary lexicographical and lexicological studies in various ways. The RSC allows the fast identification of similarly structured eponyms in large amounts of specialized text. Corpus-based lists of eponyms used as academic vocabulary and technical terms can easily be produced from this resource. The results can be sorted not only in an alphabetical order of a selected word formation component (e.g. first or second proper name or following noun), but also according to selected subject fields or journals, time periods, relevance, types of text authors, part-of-speech sequences or other linguistic and extra-linguistic criteria. Textual contexts with authentic sample sentences, information on the meaning of a term, eponymous and descriptive term variants, paraphrases, related concepts, collocations and typical syntactic environments can be extracted along with the selected types of eponyms to serve as input for dictionary databases (Ex. 8-3).

Ex. 8-3:

a) *"Here it may simply be noted that the* ***Hinshelwood-Kassel-Rice-Ramsperger treatment*** *is based on the assumption of free flow of energy between the normal modes of vibration, while the Slater treatment assumes that no such flow can occur. [...]For a given number of normal modes Slater's formula always leads to much lower rates than the* ***H.K.R.R. formula****."* (RSC, Text ID rspa_1959_0090, E. K. Gill, K. J. Laidler: The vibration and decomposition of the hydrogen peroxide molecule, 1959)

b) *"This figure cannot be reconciled with our experimentally determined activation energy of 82 kcal/mole, even by invoking* ***Rice-Ramsperger-Kassel-Hinshelwood (R.R.K.H.) theory*** *(Benson 1960) and presuming that several internal degrees of freedom can contribute energy to the decomposition process."* (RSC, Text ID rspa_1964_0106, A. G. Gaydon, G. H. Kimbell, H. B. Palmer: A shock-tube study of the kinetics of decomposition of carbon disulphide, 1964)

Users of the RSC can formulate tailored queries for specific patterns they are interested in, e.g. all terms and term variants derived from a particular name (*Golgi-Colonnier technique, Kopsch-Golgi method, Golgi-Kopsch methods*, etc.). Search results for the 1850s to the 1890s, for instance, will reveal that terms such as *Brownian movement, Brownian system, Gaussian analysis, Gaussian magnetometer, Gaussian unit, Maxwellian distribution* or *Maxwellian law* were important eponyms in this time period, while a search in the 1990s will lead to multiword constructions such as *Binley-Baginton-Wolston area* or *Gerstmann-Straussler-Scheinker disease* reflecting a change of language use tied to the ongoing specialization of disciplines.

7. Conclusions

Eponyms have not yet been addressed thoroughly and systematically for various scientific domains in previous studies and resources. Existing lexicographic products listing eponyms contribute to standardization and documentation processes, but face a particular challenge to cover these forms in a systematic way and to keep up with developments. Therefore, they can greatly benefit from insights obtained from newly available, systematically enriched specialized corpus resources such as the RSC. On the basis of the analyses in the RSC it can be confirmed that eponymization has always been an important word-formation process in academic English, not only in medical English, but also in the other disciplines and domains represented in the data. Many eponyms form groups of structurally related forms based on similar word-formation rules with varying degrees of productivity in different time periods. Personal names as term formation elements occur in combination with a variety of linguistic elements such as prefixes, adjectival or nominal suffixes and various concrete and abstract nouns. Observed regularities and trends may be considered to reflect historical strata of academic vocabulary, a change of language use tied to the ongoing specialization of individual disciplines and conventionalization processes in scientific discourse. On the one hand there is a trend towards using less morphological variety over time in eponyms, e.g. fewer adjectival suffixes and possessive forms and a more regular use of compound nouns, on the other hand previously minor word formation processes, a variety of prefixed forms as well as longer and more complex multiword patterns emerge in the more contemporary data, particularly in the mathematical, physical and engineering sciences and to a slightly lesser extent in the biological sciences. Eponymous concepts in the texts from the Royal Society journals have often been formed on the basis of non-English proper names, which suggests long-term cross-cultural contacts and exchange of ideas. Few women live on eponymously in academic discourse, particularly in the traditional fields of mainstream physics, chemistry, mathematics or astronomy. Despite various objections that have been raised against the use of eponyms and attempts by different scholars and research communities to reduce their number, eponyms are still widely used in academic papers in all disciplines represented in the RSC.

This article has discussed and traced the development of eponyms throughout the history of English scientific discourse from the early stages of the first scholarly journals published in English to contemporary scientific publications. This is both relevant for understanding linguistic patterns used in the natural sciences of the past due to the long time span the RSC covers and to understanding regularities of present-day scientific English as a high number of texts dates from the later corpus periods. There might be constraints with regard to the choice of examples and generalizations for other domains not represented in the journals' texts from the RSC, but as a large specialized dataset the corpus serves as valuable dataset of text materials with practical applications for lexicographical and lexicological studies. Resources such as specialized dictionaries can greatly benefit from a systematic use of corpus-based results for specific term formation patterns and their collocations.

Bibliography

Baayen, Harald, and Lieber, Rochelle. 1991. "Productivity and English derivation: A corpus-based study." *Linguistics*, 29(5): 801-844.

Bader, Eanas, and Shipman, Alexa R. 2015. "The Women behind the Names: Dermatology Eponyms Named after Women." *International Journal of Women's Dermatology,* 1(3): 38-41.

Bartolucci, Susan L., and Forbis, Pat. 2005. *Stedman's Medical Eponyms.* 2nd revised edition. Philadelphia: Lippincott Williams & Wilkins.

Baskett, Thomas F. 2019. *Eponyms and Names in Obstetrics and Gynecology.* Cambridge: Cambridge University Press.

Beolens, Bo, and Watkins, Michael. 2003. *Whose Bird? Men and Women Commemorated in the Common Names of Birds.* London: Christopher Helm Publishers.

Beolens, Bo, Watkins, Michael, and Grayson. Michael. 2009. *The Eponym Dictionary of Mammals.* Baltimore: Johns Hopkins University Press.

Beolens, Bo, Watkins, Michael, and Grayson. Michael. 2011. *The Eponym Dictionary of Reptiles.* Baltimore: Johns Hopkins University Press.

Beolens, Bo, Watkins, Michael, and Grayson. Michael. 2013. *The Eponym Dictionary of Amphibians.* Exeter: Pelagic Publishing.

Burdan, Franciszek, Dworzański, Wojciech, Cendrowska-Pinkosz, Monika, Burdan, Maicej, and Dworzanska, Anna. 2016. "Anatomical Eponyms—Unloved Names in Medical Terminology." *Folia morphologica,* 75 (4): 413-438.

Cabanac, Guillaume. 2014. "Extracting and quantifying eponyms in full-text articles." *Scientometrics,* 98: 1631-1645. https://doi.org/10.1007/s11192-013-1091-8.

Corpus of Academic Journal Articles (CAJA). n.d. Accessed September 06, 2020. https://www.sketchengine.eu/corpus-of-academic-journal-articles-caja/.

Degaetano-Ortlieb, Stefania, and Teich, elka. 2019. "Toward an Optimal Code for Communication: The Case of Scientific English." *Corpus Linguistics and Linguistic Theory*. (Advance online publication. https://doi.org/10.1515/cllt-2018-0088, 17/10/2019).

Dobson, Jessis. 1946. *Anatomical Eponyms: Being a Bibliographical Dictionary of those Anatomists whose Names have become Incorporated into Anatomical Nomenclature, with Definitions of the Structures to which their Names have been Attached and References to the Works in which they are Described.* London: Baillière, Tindall & Cox.

Duque-Parra, Jorge Eduardo, Llano-Idárraga, J. Oska, and Duque-Parra, Carlos Alberto. 2006. "Reflections on eponyms in neuroscience terminology." *Anatomical Record,* 289B(6): 219-224. https://doi.org/10.1002/ar.b.20121.

Fée, Antoine Laurent Apollinaire. 1858. "On the Question of whether Linnæus, in a spirit of ill-will, altered the spelling of the name of the genus Buffonia?" *Botanical Journal of the Linnean Society,* 2 (8): 183-187.

Freeman Morton S. 1997. *A New Dictionary of Eponyms.* Oxford: Oxford University Press.

Fankhauser, Peter, Knappen, Jörg, and Teich, elke. 2016. "Topical Diversification over Time in the Royal Society Corpus." *Digital Humanities 2016 Conference Abstra*cts. Kraków: Jagiellonian University & Pedagogical University, 496-500. https://dh2016.adho.org/abstracts/322.

Fankhauser, Peter, and Kupietz, Marc. n.d. *Visual Correlation for Exploring Paradigmatic Language Change*. Accessed September 06, 2020. http://corpora.ids-mannheim.de/diaviz/description.html.

Falk, Ronald J., Gross, Wolfgang L, Guillevin, Loïc, Hoffman, Gary, Jayne, David R.W., Jennette, J. Charles, Kallenberg, Cees G.M., Luqmani, Raashid, Mahr, Alfred D., Matteson, Eric L., Merkel, Peter A., Specks, Ulrich, and Watts, Richard. 2011. “Granulomatosis with Polyangiitis (Wegener’s): An Alternative Name for Wegener’s Granulomatosis.” *Journal of the American Society of Nephrology,* 22(4): 587-588. https://doi.org/10.1681/ASN.2011010081.

Fargen, Kyle M., and Hoh., Brian L. 2014. “The Debate Over Eponyms.” *Clinical Anatomy,* 27(8): 1137-1140. https://doi.org/10.1002/ca.22409.

Fischer, Stefan, Fankhauser, Peter, and Teich, Elke. 2017. “Visualization of Corpus Frequencies at Text Level.” *Proceedings of the Corpus Linguistics Conference, CL2017, Poster Session,* 25-28 July 2017, University of Birmingham, UK. https://www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2017/general/paper112.pdf.

Fischer, Stefan, Knappen, Jörg, Menzel, Katrin, and Teich, Elke. 2020. “The Royal Society Corpus 6.0—Providing 300+ Years of Scientific Writing for Humanistic Study.” Proceedings of LREC 2020, Marseille, France, May 2020, 794-802.

Fyfe, Aileen, Squazzoni, Flaminio, Torny, Didier, and Dondio, Pierpaolo. 2019. “Managing the Growth of Peer Review at the Royal Society Journals, 1865–1965.” *Science, Technology, & Human Values.* Advance online publication: https://doi.org/10.1177/0162243919862868.

Hardie, Andrew. 2012. “CQPweb—Combining Power, Flexibility and Usability in a Corpus Analysis Tool.” *International Journal of Corpus Linguistics,* 17 (3): 380-409.

Kermes, Hannah, Degaetano-Ortlieb, Stefania, Khamis, Ashraf, Knappen, Jörg, and Teich, Elke. 2016. “The Royal Society Corpus: From Uncharted Data to Corpus.” *Proceedings of LREC 2016. Portoroz, Slovenia*: 1928-1931.

Kermes, Hannah, and Teich, Elke. 2017. “Average surprisal of parts-of-speech.” *Proceedings of the Corpus Linguistics 2017 Conference,* University of Birmingham, Birmingham, UK. https://www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2017/general/paper207.pdf.

Khan, Muhammad A., and Linden, Sjef van der. 2019. “Axial Spondyloarthritis: A Better Name for an Old Disease: A Step Toward Uniform Reporting.” *ACR Open Rheumatology,* 1(5): 336-339. https://doi.org/10.1002/acr2.11044.

Kosem, Iztok. 2010. *Designing a Model for a Corpus-driven Dictionary of Academic English.* PhD thesis. Aston University, Birmingham, UK.

http://publications.aston.ac.uk/14664/1/Kosem2010_484017_3.pdf

Lalchhandama, Kholhring. 2014. "Taxonomic (r)evolution, or is it that zoologists just want to have fun?" *Science Vision,* 14: 221-233.

Letusé, La O, and Rogelio, A. 2001. *Elsevier's Dictionary of Eponyms*. Amsterdam / New York: Elsevier.

Manser, Martin. 2004. *The Chambers Dictionary of Eponyms.* Edinburgh: Chambers Harrap Publishers.

Marciano, John Bemelmans. 2009. *Anonyponymous: The Forgotten People Behind Everyday Words*. Bloomsbury USA.

McClellan, Patrick H. 2019. "Taxonomic Punchlines: Metadata in Biology." *Historical Biology: An International Journal of Paleobiology.* https://doi.org/10.1080/08912963.2019.1618293.

Menzel, Katrin. 2018. "Using Diachronic Corpora of Scientific Journal Articles for Complementing English Corpus-based Dictionaries and Lexicographical Resources for Specialized Languages." In *Proceedings of Euralex2018, Ljubljana, Slovenia*. 363-372. https://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202018/118-3-2946-1-10-20180820.pdf.

Menzel, Katrin. 2020. "Exploring English for Academic and Special Purposes through the Royal Society Corpus". In *Saarbrücken Series on Linguistics and Language Methodology*, edited by Thomas Tinnefeld. Saarbrücken: HTW Saar, 283-306.

Newmark, Peter. 1988. *A Textbook of Translation.* Hertfordshire: Prentice Hall.

Oxford English Dictionary (2000^{3}–ongoing). http://www.oed.com/.

Popescu, Floriana. 2007. "Translating Shipbuilding and Maritime Eponyms" In *Discursul Intelectual la Răspântiile Istoriei,* edited by Simona Antofi, 107-110. Galaţi: Editura Europlus.

Popescu, Floriana. 2009a. "Specialist Eponyms and General Eponyms in a Lexicographic Perspective on the English Language." *Analele Universităţii "Dunărea de Jos" din Galati, Fascicula XXIV, Lexic comun - lexic specializat*: 103-108.

Popescu, Floriana, 2009b. "Translating English Mathematical Eponyms into Romanian." *Translation Studies: Retrospective and Prospective Views,* II (5): 118-122.

Popescu, Floriana, and Maftei, Carmen. 2003. "Translating Eponyms in Physics and Applied Physics". *Colocviile Filologice Gălăţene, Editura didactică şi pedagogică, Bucureşti,* 219-233.

Popescu, Floriana, and Maftei, Carmen. 2005. "Difficulties in Translating Eponyms in Chemistry, in Cultural Perspectives." *Journal for Literary and British Cultural Studies in Romania,* no. 10: 327-334.

Quattrocchi, Umberto. 1999. *CRC World Dictionary of Plant Names: Common Names, Scientific Names, Eponyms, Synonyms, and Etymology*. Boca Raton, FL: CRC Press Inc.

Rodin, Alvin E., and. Key, Jack D. 1989. *Medicine, Literature & Eponyms: An Encyclopedia of Medical Eponyms Derived from Literary Characters*. Malabar, FL: R.E. Krieger Publishing Company.

Roth, Philippe. 2007. *Minerals First Discovered in Switzerland and Minerals Named After Swiss Individuals*. Philippe Roth, Ed.: Zurich.

Royal Society Corpus (RSC 6.0 Open). n.d. Accessed September 06, 2020. https://hdl.handle.net/21.11119/0000-0004-8E37-F.

Stigler, Stephen M. 1980. "Stigler's Law of Eponymy." *Transactions of the New York Academy of Sciences* 1980 (Festschrift for Robert K. Merton), 39: 147-57. https://doi:10.1111/j.2164-0947.1980.tb02775.x.

Van Tassel, Sarah H., Segal, Kira L., Ciralsky, Jessica B., and Sun, Grace. 2018. "Eponymous Women in Ophthalmology: Syndromes with Prominent Eye Manifestations named after Female Physicians." *Eye,* 32 (8): 1293-1295.

Whitworth, Judith Ann. 2007. "Should Eponyms be Abandoned? No." *BMJ* 335: 425.

WHO. "Best Practices for Naming of New Human Infectious Diseases". World Health Organisation, Geneva2015. Accessed September 06, 2020. https://www.who.int/topics/infectious_diseases/naming-new-diseases/en.

Woywodt, Alexander, and Matteson, Eric. 2007. "Should eponyms be abandoned? Yes." *BMJ* 335: 424.

CHAPTER NINE

GERMAN PARONYM DICTIONARIES: THE SHIFT FROM PRESCRIPTIVE PRINT EDITIONS TO ELECTRONIC CORPUS-BASED RESOURCES

PETRA STORJOHANN

Abstract

This article sketches the development of paronym dictionaries in German. These dictionaries document and describe commonly confused words which cause uncertainties because they are similar in sound, spelling and/or meaning (e.g. *effektiv*/*effizient*, *sportlich*/*sportiv*). First, an overview of existing reference guides is provided, covering different traditions. Numerous lemma lists have been collected for pedagogical purposes and there has always been an interest in the lexicological treatment of paronyms. However, only a handful of dictionaries covering commonly confused pairs and a small number of genuine paronym dictionaries have ever been compiled. I will focus on lexicographic endeavours, including Wustmann (1891), Müller (1973) and Pollmann and Wolk (2001). Secondly, I will shed light on the differences in descriptions in these dictionaries. This includes how prescriptive approaches have been replaced over time by empirical descriptive accounts and how dictionaries have moved away from restricted, static hardback editions towards dynamic e-dictionaries. Finally, an e-dictionary, “Paronyme—Dynamisch im Kontrast”, is presented with contrastive and flexible two-level consultation views. Its three key elements are its corpus-based foundation, the implementation of meta-lexicographic requirements and a consideration of users’ interests. This dictionary has implemented a user-friendly and dynamic interface and it records conventionalized patterns and preferences in authentic communication.

1. Introduction

In everyday conversations, speakers regularly come across words that they misuse or confuse. There are many reasons why people mix up words or why a specific word evokes another morphologically or semantically related or unrelated word. Sometimes, a lexical item is simply incorrectly stored in the mental lexicon or incorrectly associated with another term. In other cases, two terms are very similar to each other with regard to spelling, sound or meaning. We encounter commonly confused terms that are involuntarily linked in every language.

> We say one word when we mean another, half-comprehend or misunderstand words, and encounter unfamiliar and 'hard' words daily. In short, we confuse words. (Room 1979, 1)

Today, rather than consulting dictionaries, or possibly because users encounter dictionary problems, people turn to online forums and ask the language community for advice. Questions such as "Is there a difference between the terms *suffragette and suffragist,* or do they mean the same?" or "What is the difference between *authoritative* and *authoritarian*?" are a clear indication of users experiencing linguistic issues that they want to resolve. In these forums, people share their knowledge and expertise but also their opinions and prejudices.

With respect to German, confusing words are, in fact, not systematically documented in standard dictionaries although they have been of interest to a small number of lexicographers for over a hundred years now. The first dictionary by Wustmann (1891), rather random and limited in scope, followed a prescriptive tone, pointing to the correct usage of alternative plural forms. Specialized dictionaries were subsequently written by larger publishing companies such as Duden (Müller 1973) and PONS (Pollmann and Wolk 2010), appearing as new lexicographic authorities trying to disentangle all types of frequently confused words. This included words which have been misused or confused both by learners of German and by native speakers. A dictionary solely dedicated to genuine paronyms (see section 2, below), however, has never been systematically compiled, as Hausmann (1990) rightly points out.

With the dawn of corpus linguistics and the development of Internet technologies, lexicographic traditions have been exposed to major changes over the last twenty-five years. The German lexicographic landscape has only gradually adjusted to new editorial processes, to changes in users' interests and to new design options (cf. Müller Spitzer 2014). The following sections will outline how confusables have been treated lexicographically over the past hundred years and why it has been necessary to develop both new lexical approaches to the phenomenon and an alternative, above all empirically sound and reliable, reference guide to contemporary German paronyms. This article will survey the more recent development of German paronym dictionaries and look at a new online dictionary in more detail. In particular, two insightful works that were published as print volumes prior to the rise of corpus-based electronic lexicography will be reviewed. This will be followed by a detailed description of the latest addition to dictionaries covering confusables, primarily paronyms. This reference guide is also the first online dictionary of German confusable pairs which makes extensive use of corpus-guided, descriptive and synchronic principles and which offers dynamic information on demand for contrastive consultation.

2. Paronyms and other commonly confused terms

2.1 Confusables in language use

In every language, there are terms that are commonly confused lexical items. Often these are words that are separated by just one or two letters, and sometimes also prefixes or suffixes. With respect to German, speakers confuse such words for different reasons, such as close semantic meaning, e.g. *kalt/kühl* (*cold/crisp*) (near-synonyms), similarity or identity of spelling, e.g. der *Band*/das *Band*/die *Band* (*volume/ribbon/band*) (homographs) or because the words are identical in sound, such as (*Leib/Laib*) (*body/loaf*) (homophones). These rather prototypical cases do not account for a full classification of commonly confused terms; the reasons for them and the effects of the confusion are complex (cf. Klein 2018). As well as lexical confusion, confusables can also be the result of grammatical confusion, such as difficulties arising from varying inflection, the usage of neologisms and loan words and uncertainties surrounding word formation patterns, congruence and variable genders of nouns, to name but a few.

Paronyms are a specific group of confusables: they are lexical items where one word resembles another in spelling or pronunciation and they sometimes also have a similar or associated meaning. They are words that have a lexical, semantic and sometimes also etymological affinity with another. These can be pairs, e.g. *effektiv/effizient* (*effective/efficient*), *autoritär/autoritativ* (*authoritarian/authoritative*), *sportlich/sportiv* (*sporty/athletic*), but also larger groups, e.g. *provokant/provozierend/provokativ/provokatorisch* (*provocative, provoking*) (cf. Room 1979, 3).

2.2 Paronyms in linguistics

The phenomenon of paronymy (easily confused words) has not attracted much attention, either from a corpus linguistic or from a cognitive linguistic perspective. The most comprehensive theoretical approach so far is offered by Lăzărescu (1999). His model treats paronyms from a structuralist point of view, accounting for language as a formal and logical system, and is not based on empirical evidence in real communicative situations. Looking at this relation from a language learner's perspective and with approaches used in translation studies, Lăzărescu developed an elaborate model based on strict formal criteria, primarily word formation and syntax. He aimed to establish clear-cut boundaries between paronymy and other phenomena of lexical confusion, such as homographs, homophones, lexical alternatives and false friends, etc. Still, fundamentally his model was not based on large amounts of empirical evidence of language use and consisted of the following main categories: phonetic-orthographic aspects (*Föhn*/*Fön*), morphological aspects (*Kinderliebe*/*Kindesliebe*; *seiden*/*seidig*), syntactic aspects (*nachts*/*nächtlich*; *schuld*/*schuldig*) and stylistic aspects (*essen*/*fressen*).

With his traditional classification Lăzărescu opened up the study of confusables but we still lack a definition of the phenomenon from a usage-based perspective incorporating cognitive aspects. Any definition of paronymy can thus only be an approximation of a rather complex phenomenon which is still widely under-researched. In this article, paronymy provisionally refers to two or more lexical items from the same language and the same word class that are similar in sound, writing and/or meaning. These items often have the same etymological root but differ with respect to endings or prefixes. They often have semantic overlappings or denote similar concepts but sometimes they do not. Due to their similarities, they run a higher risk of being confused and misused in certain contexts. Consequently, some of Lăzărescu's categories are implied in this broad and vague definition, others are excluded, also because they are more relevant from a language learners' view (e.g. stylistic aspects) or they don't cause confusion in oral communication (e.g. phonetic-orthographic aspects).

From what we know so far, corpus studies have revealed that such categorizations are incompatible with language in use and with cognitive elements relating to confusion and misuse (cf. Mell, Schnörch and Storjohann 2019). Today, we are concerned with an empirically sound, usage-guided investigation of commonly confused words based on large corpus data. Different linguistic and extra-linguistic aspects have an impact on the potential lexical misuse of words. From a corpus-based and language-oriented perspective, the criteria for distinguishing types of paronyms mainly concern communicative functions, discourse affiliation and diverse semantic properties. Currently, investigations also focus on research into paronymy as a complex lexical-conceptual phenomenon, aiming to develop an empirically driven classification of paronyms using diverse genres of language evidence and including written and spoken texts (cf. Mell, Schnörch and Storjohann 2019). So far, we have gained valuable insights into functions in specific contextual instances, communicative functions, thematic domains, discourse and style, text types and degrees of semantic similarity or contrast between easily confused words. Furthermore, speakers' attitudes can be expressed through their choice of paronyms while encyclopaedic knowledge and cultural experience also play a key role in the use and interpretation of specific discourse-bound word pairs. These influential elements can be detected through collocations and grammatical constructions in context. They are more or less conveyed meta-linguistically and are therefore explicit in written communication.

A classic case of paronymy is caused by similarity in sound and spelling as in *ethisch*/*ethnisch* (*ethical*/*ethnical*). Semantically, they do not refer to similar concepts; there is no semantic overlap at all. They are confused in both directions due to their formal likeness. As a result, linguistic errors appear which can cause communicative problems and misinterpretations.

A second group comprises paronyms that are similar in meaning but exhibit slight semantic-pragmatic nuances in terms of preferences for corresponding discourse. The case of *sportlich*/*sportiv* (*sporty*/*athletic*) shows how both adjectives depict a person as being athletic, healthy and fit. Whereas *sportlich* often relates to the appearance of a person and is preferred in the context of crime and police reports in order to describe offenders, *sportiv* more often occurs in the context of sport and health issues. As both items can describe appearances, their meanings and references are identical. Nonetheless, they differ with respect to their function of reflecting the appropriate discourse. Confusing these terms, however, does not cause communicative misunderstanding.

A third category concerns opposites such as *konkav*/*konvex* (*concave*/*convex*). These have identical collocates which are characterized as having a specific type of curvature (*Spiegel*, *Augen*, *Linse*, *Scheibe*—*mirror*, *eyes*, *lens*, *disc*/*plate*). Curvature relates to a bipolar concept describing an entity having an outline that curves inwards or outwards. Each way is lexically realized by its corresponding paronym. Lexical confusion is motivated by extra-linguistic circumstances and not caused by semantic uncertainties. In this case, speakers confuse the concept and its lexical realization. Often they are aware of their own linguistic doubts, and mnemonic tricks are common to avoid misuse.

A fourth class consists of semantic alternatives without differences or semantic nuances. The adjectives *patriarchisch*/*patriarchal*/*patriarchalisch* (*patriarchal*) and *provokant*/*provozierend*/*provokativ*/*provokatorisch* (*provocative*) cause linguistic doubts because speakers expect differences between the items. They often do not believe that lexical alternatives derived from the same noun can co-exist and can also denote the same concept.

Moreover, there are technical terms which are used synonymously in public discourse but are strictly defined in their specific technical discourse, e.g. *Parodontose*/*Parodontitis* (*paradontosis/periodontitis*) in medicine and *antisozial*/*asozial*/*unsozial*/*dissozial* (*antisocial*/*asocial*/*unsocial*/*dissocial*) in psychology.

A number of pairs can be detected which behave differently in official language use (e.g. the discourse of administration). To avoid political incorrectness, to dispense with stigmatized items and/or to adapt to new social or cultural circumstances, new terms are officially introduced, causing uncertainties in public discourse, such as *unehelich*/*nichtehelich*/*außerehelich* (*misbegotten*/*illegitimate*/*out of wedlock*).

By far the largest group of paronyms concerns lexical items which share a number of contexts in which they can be used synonymously and a number in which this is not the case (*effektiv*/*effizient*, *sensitiv*/*sensibel*). The degree of similarity and difference can vary largely, allowing for further sub-classifications.

Overall, defining and classifying paronyms is a complex matter. Paronymy is not a lexical relation but a dynamic conceptual relation with cognitive implications which are visible on a linguistic level. In order to develop a full model, the identification of communicative functions and influences on lexical confusions is necessary. The effects of lexical misuse open up a number of questions concerning misunderstanding or semantic change. A description of the components involved in lexical confusion allows for a better understanding of paronymy in authentic use in all its complexity and with all its facets. How paronyms have been treated lexicographically and how traditions have shifted from prescriptive to descriptive accounts based on different lexical data, methods and presentations will be looked at in more detail in the following sections.

3. Paving the way to document confusables

For a long time, German was considered to be the language of the uneducated, a language inferior to the more prestigious Latin and Greek. Both of these were the gateways to contemporary scholarship, and the specifics of German only came up in contrast to Latin or Greek (cf. Klein 2018, 43-44). Due to a lack of interest in the study of the system of the German language in general, interest in cases of lexical uncertainty arose no earlier than the seventeenth/eighteenth century. This was the time when it was increasingly felt that there was a need to prescribe linguistic standards. With a particular focus on grammatical variation of different types, renowned authors and scholars faced a growing demand to select the allegedly correct form.[1] Some favored the language of Luther; others preferred certain literary traditions or dialectal variations. Changes of gender, tense and number formed numerous cases of variables of the same lexical item. Among the first attempts were systematic studies of syntactic or spelling variations of lexical forms and the categorization of these according to their origins and affiliations.[2] Particularly, between the late eighteenth and the early nineteenth centuries, a number of authors, scholars, philologists and educationists increasingly dedicated their time to linguistic issues. It was at this time that the first German-language-based studies and dictionaries were published, describing varying inflectional forms, gender variation and in particular spelling variation and how to regulate these. Hanks's (2013) observation with respect to English also holds true for German, at least in parts:

> These standards were based on the ill-defended assumptions that earlier forms of a language are somehow more 'correct' than contemporary forms and that etymology guarantees meaning (Hanks 2013, 514).

Among these are, for example, forerunners such as Gottsched (1762) with his book "Deutsche Sprachkunst: Ein orthographisches Verzeichniß gewisser zweifelhafter Wörter", Adelung (1788) with "Vollständige Anweisung zur Deutschen Orthographie, nebst einem kleinen Wörterbuch für die Aussprache, Orthographie, Biegung und Ableitung" and Sanders's (1900) "Wörterbuch der Hauptschwierigkeiten in der deutschen Sprache". These authoritative works were general accounts of the German language which only gradually became the subject of general linguistic studies. The aims of these puristic guidebooks were twofold: first, to educate, improve and enhance language skills in order to enable speakers to choose the correct alternative by avoiding the stigmatized, impure or wrong option; and secondly, to deal with irritating cases and eradicate 'ugly', 'troublesome'

[1] Linguistic correctness and overcoming grammatical variation was, for example, demanded by Leibniz (1679) in order that German could be of equal status to Latin, cf. Pörksen (1983).
[2] See for example Ölinger (1573).

linguistic phenomena. They were driven by their aspirations to establish norms in times that were associated with the assumed decline of linguistic knowledge due to a lack of education and a growing deterioration in linguistic competence. With their dictionaries, scholars wanted to rescue linguistic norms from the language decay caused by the omnipresent and ever-growing industrialization. Vernacular and colloquial forms, for example, were considered impure and improper. Opinionated editors clearly anticipated cultural snobbery and insisted on the lexical forms which had been documented in respectable and reputable linguistic works (usually the ones offered by exemplary classical authors). It is an approach which makes the single opinion of a scholar the presiding authority and the final decision about what is right or wrong.

With growing literacy, progressing (written) standardization and supra-regionality of German, a number of quick-reference lists and authoritative linguistic collections were published to present and define 'correct' words and to impose linguistic order, mainly for educational purposes.[3] By the end of the nineteenth century, High German had been effectively pedagogically promoted as a standard and established itself as a prestigious social norm. Although the dictionaries mentioned above are not paronym dictionaries, they clearly paved the way for instructive linguistic resources, which later became an established lexicographic tradition in the nineteenth and twentieth centuries. They place considerable emphasis on linguistic problems, mistakes and deficiencies, as most of the titles suggest.[4]

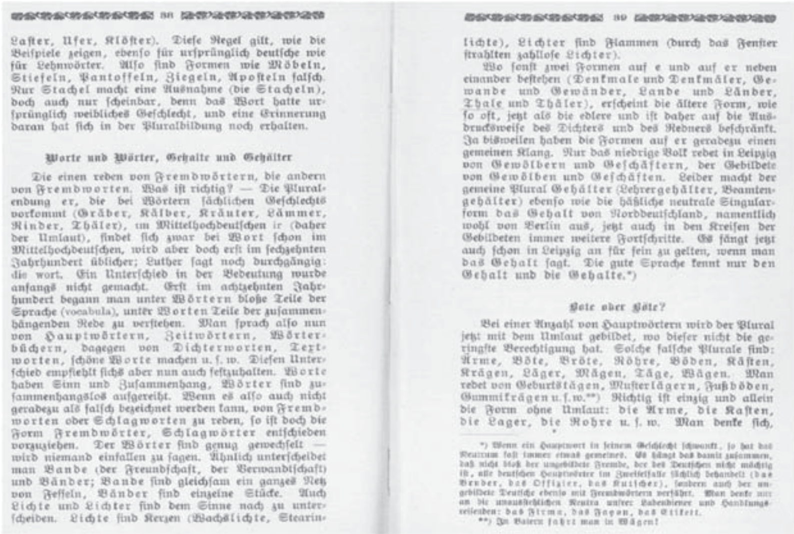
Originally published in 1891, Gustav Wustmann's reference guide *Allerhand Sprachdummheiten: Kleine deutsche Grammatik des Zweifelhaften, des Falschen und des Häßlichen*[5] is, in fact, the first practical guide book which explicitly sets out a number of specific easily confused word pairs. Its value resides in a small chapter which contains entries for pairs of alternative plural forms such as *Worte*/*Wörter* (*words*) or *Gehalte*/*Gehälter* (*income*) (see Fig. 9-1).

---

[3] Klein (2018) also outlines the politically motivated reasons for the standardization of German. With the foundation of Germany as a nation-state in 1871, there was a strong desire for the establishment of a uniform, nationwide linguistic standard, starting with a conference on standardized spelling in 1901.

[4] For a detailed list of new prescriptive resources of the nineteenth century concerning imperfect language use, see Klein (2018, 63-66).

[5] This title translates as "All sorts of linguistic stupidities: concise German grammar of the doubtful, the wrong and the ugly". It continued to be edited until 1966.



38

Laster, Ufer, Klöster). Diese Regel gilt, wie die Beispiele zeigen, ebenso für ursprünglich deutsche wie für Lehnwörter. Also sind Formen wie Möbeln, Stiefeln, Pantoffeln, Ziegeln, Aposteln falsch. Nur Stachel macht eine Ausnahme (die Stacheln), doch auch nur scheinbar, denn das Wort hatte ursprünglich weibliches Geschlecht, und eine Erinnerung daran hat sich in der Pluralbildung noch erhalten.

Worte und Wörter, Gehalte und Gehälter

Die einen reden von Fremdwörtern, die andern von Fremdworten. Was ist richtig? — Die Pluralendung er, die bei Wörtern sächlichen Geschlechts vorkommt (Gräber, Kälber, Kräuter, Lämmer, Rinder, Thäler), im Mittelhochdeutschen ir (daher der Umlaut), findet sich zwar bei Wort schon im Mittelhochdeutschen, wird aber doch erst im sechzehnten Jahrhundert üblicher; Luther sagt noch durchgängig: die wort. Ein Unterschied in der Bedeutung wurde anfangs nicht gemacht. Erst im achtzehnten Jahrhundert begann man unter Wörtern bloße Teile der Sprache (vocabula), unter Worten Teile der zusammenhängenden Rede zu verstehen. Man sprach also nun von Hauptwörtern, Zeitwörtern, Wörterbüchern, dagegen von Dichterworten, Textworten, schöne Worte machen u. s. w. Diesen Unterschied empfiehlt sichs aber nun auch festzuhalten. Worte haben Sinn und Zusammenhang, Wörter sind zusammenhangslos aufgereiht. Wenn es also auch nicht geradezu als falsch bezeichnet werden kann, von Fremdworten oder Schlagworten zu reden, so ist doch die Form Fremdwörter, Schlagwörter entschieden vorzuziehen. Der Wörter sind genug gewechselt — wird niemand einfallen zu sagen. Ähnlich unterscheidet man Bande (der Freundschaft, der Verwandtschaft) und Bänder; Bande sind gleichsam ein ganzes Netz von Fesseln, Bänder sind einzelne Stücke. Auch Lichte und Lichter sind dem Sinne nach zu unterscheiden. Lichte sind Kerzen (Wachslichte, Stearin-

39

lichte), Lichter sind Flammen (durch das Fenster strahlten zahllose Lichter).

Wo sonst zwei Formen auf e und auf er neben einander bestehen (Denkmale und Denkmäler, Gewande und Gewänder, Lande und Länder, Thale und Thäler), erscheint die ältere Form, wie so oft, jetzt als die edlere und ist daher auf die Ausdrucksweise des Dichters und des Redners beschränkt. Ja bisweilen haben die Formen auf er geradezu einen gemeinen Klang. Nur das niedrige Volk redet in Leipzig von Gewölbern und Geschäftern, der Gebildete von Gewölben und Geschäften. Leider macht der gemeine Plural Gehälter (Lehrergehälter, Beamtengehälter) ebenso wie die häßliche neutrale Singularform das Gehalt von Norddeutschland, namentlich wohl von Berlin aus, jetzt auch in den Kreisen der Gebildeten immer weitere Fortschritte. Es fängt jetzt auch schon in Leipzig an für fein zu gelten, wenn man das Gehalt sagt. Die gute Sprache kennt nur den Gehalt und die Gehalte.*)

Bote oder Böte?

Bei einer Anzahl von Hauptwörtern wird der Plural jetzt mit dem Umlaut gebildet, wo dieser nicht die geringste Berechtigung hat. Solche falsche Plurale sind: Ärme, Böte, Bröte, Röhre, Böden, Kästen, Krägen, Läger, Mägen, Täge, Wägen. Man redet von Geburtstägen, Musterlägern, Fußböden, Gummikrägen u. s. w.**) Richtig ist einzig und allein die Form ohne Umlaut: die Arme, die Kasten, die Lager, die Rohre u. s. w. Man denke sich,

*) Wenn ein Hauptwort in seinem Geschlecht schwankt, so hat das Neutrum fast immer etwas gemeines. Es hängt das damit zusammen, daß nicht bloß der ungebildete Fremde, der des Deutschen nicht mächtig ist, alle deutschen Hauptwörter im Zweifelsfalle sächlich behandelt (das Bruder, das Offizier, das Kutscher), sondern auch der ungebildete Deutsche ebenso mit Fremdwörtern verfährt. Man denke nur an die unausstehlichen Neutra unserer Ladendiener und Handlungsreisenden: das Firma, das Fagon, das Etikett.

**) In Baiern fährt man in Wägen!

Fig. 9-1: Chapter on variable plural forms *Worte*/*Wörter*, *Gehalte*/*Gehälter* in Wustmann (1891).

The corresponding entry has a variety of information, all in narrative form. It elucidates varying plural forms of *Wort* (*word*) and their etymological origins, including their Middle High German forms. The entry informs the reader that there was no difference in meaning before the eighteenth century: only later did speakers distinguish between the two plurals. While *Wörter* refers to single lexical items, the term *Worte* designates a coherent speech. There then follows clear advice for maintaining this difference. As Wustmann (1891, 38) proclaims: "Diesen Unterschied emphielt sichs aber nun auch festzuhalten".[6] At the same time, with respect to corresponding compounds, he states that it is not wrong to use *Fremdworte* (*loan words*) instead of *Fremdwörter*, but it is clearly not to be preferred. Ruthlessly making recommendations based not on linguistic insights but rather on popular views, Wustmann refers to the different preferences among the lower and educated classes, the noble vs. the ugly, as well as the common nature of the two forms and their 'mean sounds'. Wustmann's aim was to educate the careless German language user in how to use these words correctly. With respect to varying forms, his book offers clear decisions about what is right and wrong. Some linguistic cases were

[6] It translates as: "This difference mus be maintained".

clearly labelled as ‘wrong’, ‘flaws’ or ‘common errors’, and it was usually the oldest or allegedly aesthetic lexical version that he was inclined to characterize as appropriate. His descriptions are full of judgmental positions on linguistic variation (for more, see Klein 2018, 67), as was typical for that time. Being a conservative guide with a strictly prescriptive style of explanation, with successive revisions, particularly the tenth edition in 1935, it evolved and contained a growing number of paronym examples, paving the way for activities fundamentally dedicated to cases of lexical confusion in language use. It is largely in this spirit that his pioneering lexicographic work is appreciated today, as it had a profound effect on future specialized lexicography of confusable pairs (cf. Lăzărescu 1999; Klein 2018).

4. Introspective print editions

4.1 Duden “Leicht verwechselbare Wörter” (Müller 1973)

In the twentieth century, the phenomenon of commonly confused items including paronyms received more attention. Duden (Müller 1973) published the first specialized reference guide dedicated to different types of lexical confusables. At 344 pages, its paperback volume “Leicht verwechselbare Wörter” (easily confusable words) (see Fig. 9-2) constitutes a milestone for two reasons. First, it was the first attempt to compile a German inventory of confusables. Secondly and equally importantly, it is a large introductory survey elucidating the subject of misused words in more detail, including their types, causes and effects in communication. With his work, Müller (1973) hoped to spark new interest in lexical misuse and confusion. As such, we encounter a lexicological perspective of the phenomenon, gradually raising an awareness of the widespread existence of linguistic doubt and lexical confusion/misuse. In accordance with existing lexicographic approaches, its treatment of meaning and usage is normative and restrictive in nature. In the preface, the author stresses that the confusion between two words is a mistake caused by a lack of either concentration or knowledge (cf. Müller 1973, 8). Clearly, this book aims to educate native speakers of German as much as learners of German who need to be equipped with the correct and necessary linguistic knowledge. Generally, it provides tips to pre-empt some of the most frequent errors; its aim is to assist users in finding the correct lexical items for correct contexts so they can distinguish between right and wrong.

From a lexicographic view, we can identify the evident problem of randomness of headwords, as this reference book ranges from cases of semantic interference between words that are similar in sound, e.g. *Biographie*/*Bibliographie* (*biography*/*bibliography*) to homonyms, e.g. *das Schild*/*der Schild* (*sign*/*shield*). There are also selected alternative inflected forms, such as participle forms, e.g. *gesinnt*/*gesonnen* (*minded*) and numerous technical terms, e.g. *Phrase*/*Phraseologie* (*phrase*/*phraseology*). Most pairs or groups, however, involve paronym cases such as *antik*/*antikisch*/*antiquiert*/*antiquarisch* (*antique*/*ancient*/*antiquated*/*antiquarian*). Some obscure entries, e.g. *sensibel/sensitiv/sentimental/sentimentalisch/sensuell/sensualistisch/sensorisch/sensoriell/senil*, consist of up to nine adjectives (see Fig. 9-2).

sensibel/sensitiv/sentimental/sentimentalisch/sensuell/sensualistisch/sensorisch/sensoriell/senil

Als **sensibel** wird jemand bezeichnet, der feiner seelischer Empfindungen fähig ist. Ein sensibler Mensch ist feinfühlig, einfühlsam, empfindsam, zartfühlend und alles andere als robust:

> es gäbe Beispiele, daß sensible Kinder sich erst nach Ablauf einiger Wochen eingewöhnten (Jens, Mann 53); Sensible Naturen werden ja schon durch den Auspufflärm halb verrückt gemacht, während sich die Robusteren. . . nicht einmal durch das Abfeuern einer Kanone stören lassen (Menzel, Herren 82); seit ich weiß, wie sensibel sie ist, . . . habe ich nie wieder gefragt (Frisch, Gantenbein 326).

Im übertragenen Gebrauch bedeutet *sensibel* soviel wie *Behutsamkeit und Vorsicht im Umgang erfordernd oder zeigend:*

> man müsse mit einer so sensiblen Materie wie mit Nachrichten vorsichtig umgehen (Bundestag 189/1968, 10243); beeindruckend vor allem die sensible Führung der Kamera, die Szenen und Gesichter in ihren charakteristischen Zügen herauszuschälen weiß (MM 4./5.6.66).

Eine Art Steigerung von *sensibel* enthält das Adjektiv **sensitiv**, das *übermäßig empfindsam, empfindungsfähig* bedeutet. Der sensitive Mensch mit seiner gesteigerten Empfindsamkeit und Überempfindlichkeit sowie seiner starken Verletzbarkeit des Selbstwertes neigt zu heftiger Nerven- und Gefühlserregbarkeit und zu Neurosen in Verbindung mit Selbsterniedrigung, Selbstbeschuldigung und Selbstquälerei:

> Der Typus des sensitiven Psychopathen. . . ist inzwischen in die meisten modernen Lehrbuchdarstellungen übergegangen (Kretschmer, Beziehungswahn 9); Er hat ein Organ für die sensitive Sphäre, für Gemütsspannungen und Empfindungsweisen (Deschner, Talente 55); Er . . . sang in sich hinein, denn sein Befinden war musikalisch und sensitiv (Th. Mann, Zauberberg 197).

Im Englischen haben die lautlich entsprechenden Wörter ganz andere Bedeutungen, was sich manchmal auch in deutschen Texten widerspiegelt. Das englische *sensitive* entspricht der Bedeutung des deutschen Adjektivs *sensibel*, während das englische *sensible* soviel wie *vernünftig* bedeutet.

Während mit dem Adjektiv *sensibel* auf die innere Erlebnisfähigkeit eines Menschen hingewiesen wird, kennzeichnet **sentimental** die auch nach außen sichtbar werdende Hingabe an ein Gefühl oder an eine seelische Stimmung, die oft an Rührseligkeit grenzt. Ein sentimentaler Mensch ist gefühlsselig, schwärmerisch, also ein wenig zu gefühlsbetont und romantisch:

> der Jazz-Sänger sang ein sentimentales Lied (Koeppen, Rußland 112); Ich saß in der Bar – Reisebekanntschaften! Ich wurde sentimental (Frisch, Homo 124); Eine schmale. . . Kette aus Golddoublé hatte das sterbende Mädchen um den rechten Fuß getragen – eine dieser Albernheiten, zu denen man nur fähig war, wenn man jung, sentimental und ohne Geschmack war (Remarque, Triomphe 20).

Das Adjektiv **sentimentalisch** wird nur selten gebraucht. Ursprünglich waren *sentimental* und *sentimentalisch* gleichbedeutend und die Endung -isch bedeutete nur eine Angleichung an das deutsche Wortbildungssystem, wie es auch sonst öfter der Fall ist (perfektiv/perfektivisch). Im Unterschied zu *sentimental*, das eine innere Qualität, Beschaffenheit charakterisiert, gibt *sentimentalisch* jedoch an, in welcher Art, in welchem Stil sich etwas vollzieht (vgl. antikisch, genialisch, idealisch):

> Das sentimentalische Fernweh ist eine romantische Kategorie (Enzensberger, Einzelheiten I 186).

Im literarischen Gebrauch hat sich – so bei Schiller – eine besondere Bedeutung dieses Wortes entwickelt, und zwar wurde es als Gegensatz zu *naiv* gebraucht. Unter sentimentalischer Dichtung versteht Schiller eine Dichtung, die bemüht ist, die durch Kultur und Zivilisation verlorengegangene ursprüngliche Natürlichkeit durch Refle-

232

Fig. 9-2: Extract from Duden's (Müller 1973) entry *sensibel/sensitiv/sentimental/sentimentalisch/sensuell…*

Its entry structure is straightforward, implementing a typical traditional lexicographic architecture. This encompasses a narrative definition with synonyms, and each paraphrase is supported by an illustrative citation from literature or occasionally by an example from a newspaper. Moreover, in an entry, each lemma is described one after another with information on how to avoid common mistakes. Strikingly, this dictionary adopts a different policy in relation to its defining style. Meanings are not strictly and formally defined according to genus and differentia (the Aristotelian-Leibnizian principles of definition). Instead, they are paraphrased in full sentences and the headword's senses are rather typified. In addition, a conceptual reference is implemented in the definition, characterizing the subject that is typically associated with the adjective in question, in this case *sensitiv*. The entry can also include explanations of metaphoric shifts or false friends. The contrast between two or more words is highlighted by pointing out their semantic or stylistic differences. With polylexematic entries, an explicit description of the contrast between all individual headwords is usually absent.

Looking at the example, we have a dictionary that is quite modern in its approach to data. Although the corpus (a citation collection) is not representative from today's perspective, it is not used as the empirical base to detect headwords systematically nor is it used to examine them semantically. Nonetheless, all entries are consistently equipped with authentic quotations instead of invented examples. It is quotations from up-market sources—often classic or renowned fiction—that illustrate the explanations. Even though it is an explicit endeavour to represent a linguistic reality, the semantic descriptions themselves remain introspective treatments of large groups of lexemes which brings up a number of disadvantages, in terms of presentation and content. As Veith (1977) points out:

> Der Erklärungsversuch des Verfassers ist die „innersprachliche" Interferenz, die allerdings wenig fundiert wird und so eine anfechtbare Grundlage bildet für die Zusammenstellung der Lemmata (Veith 1977, 225).[7]

As well as the questionable choice of some headwords, some semantic deviations are far-fetched or misleading. For example, *sensitiv* is described as a type of semantic comparative of *sensibel*, a rather implausible remark.

[7] "The authors' attempt to explain this is 'internal' interference, which, however, has little foundation and thus forms a contestable basis for the compilation of the headwords" (translated by the author).

Another disadvantage is the omission of senses which leads to narrow views on the use of the words. Here, *sensitiv* and *sensibel* are only used to refer to people with specific characteristics.[8] Particularly with regard to polysemous items, there appears to be only a frequent or prototypical use, stripping down a meaning spectrum to a singular usage event and ignoring further regular and common occurrences. Essentially, most meanings are not captured at all and any semantic similarities between words are not listed.

As demonstrated in Fig. 9-2, a distinctly straightforward structure does not seem feasible for such large entries, nor are the terms in question adequately contrasted with each other individually. From today's perspective, this dictionary is outdated in many ways. Nonetheless, it is the chief source of information on confusables of its time, and it is a stepping-stone to a more theoretical and practical linguistic enterprise in terms of commonly confused words.

4.2 PONS "Leicht verwechselbare Wörter" (Pollmann and Wolk 2010)

More recently, and at a time when German corpora and corpus tools had been widely available for over a decade, PONS (Pollmann and Wolk 2010) published a new dictionary of lexical confusables titled "Leicht verwechselbare Wörter". This reference guide was compiled by language tutors who teach German as a foreign language to students at the university level, so their perspective is a language learner's view of confusing words. This volume reflects some discrepancies within the field of lexicography. Most German lexicographers were already using various online databases at least to locate and confirm quotations or even for systematic searches of lexemes and language patterns. It is therefore surprising to start a new lexicographic project without considering such databases and established corpus procedures.

Their approach to the detection of headwords and to documentation is ambivalent. Their database (archival material) and editorial methods are neither representative nor systematic and their objective is not the analysis and interpretation of regularity, significance and typically recurrent patterns of actual language use. The headwords of this dictionary represent a more or less random aggregation of lexical pairs that the authors have come across over a number of years. By contrast, the guide does not follow a prescriptive approach but rather uses a non-conformist defining style without strict rules

[8] For a full description of *sensitiv*/*sensibel*, see Storjohann (2015).

about the use of certain words.[9] Although most definitions lack information related to conceptual references, they show contextual conditions in which the word in question is used. This means the strength of this book lies in its descriptive tone. Instead of unintelligible definitions that follow certain conservative formulae, the meanings described here reflect how to use words in context. Readers are being informed rather than corrected; they are invited to browse and to be entertained.

However, as a print volume, its limits are, above all, its space constraints. This rather small dictionary contains roughly 1000 headwords. Among them are semantically reversive pairs such as *lehren*/*lernen* (*teach*/*learn*), homophones such as *Nachname*/*Nachnahme* (*surname*/*cash-on-delivery*), near-synonyms such as *Krebs*/*Tumor* (*cancer*/*tumor*), regional alternatives such as *Berliner/Pfannkuchen (jam donut)*, rhyming pairs such as *loyal*/*royal* and also a diverse spectrum of confusable words[10] with some genuine paronyms (*elektrisch*/*elektronisch*, *technologisch*/*technisch*) among them (see Fig. 9-3).

Looking at the example of *elektrisch*/*elektronisch*, two highly polysemous words, the entry focusses only on semantic differences in terms of their operating means of technical devices, energy and control elements such as transistors.

elektrisch/elektronisch

- Sprechen wir von ***elektrisch***, so sprechen wir vom Nutzbarmachen der Elektrizität, also zum Beispiel der Stromerzeugung. Das Ziel ist primär die Energieübertragung.

- **Elektronisch** hingegen umschreibt die technische Ausnutzung dieser Elektronensteuerung, zum Beispiel in Fernsehen, Radios oder Haushaltsgeräten. Elektronik dient damit also vorrangig der Informationsübertragung.

Fig. 9-3: PONS entry of *elektrisch/elektronisch.*

The translation of the above entry follows (see Fig.9-3a):

---

[9] This approach is similar to an earlier English dictionary, *Room's Dictionary of Confusables* (1979).

[10] It also contains rather strange pairs such as *Parade*/*Pomade* (*parade*/*pomade*). The items are neither etymologically related, thus showing no morphological overlap, nor do they have similarities in terms of semantics or sound." They also do not occur in relevant language forums discussing confusables.

- When we speak of *electric*, we speak of the utilization of electricity, for example, the generation of electricity. The ultimate goal is the transfer of energy.
- *Electronic*, however, describes the technical exploitation of this electron control, e.g. as in televisions, radios and household appliances. Electronics thus primarily serves to transfer information.

Fig. 9-3a: Translation: PONS entry of *elektrisch/elektronisch*

In actual language use, however, there are a lot of commonalities. For instance, *music* can be *electric* or *electronic*, both referring to the same synthetic sounds without marking a difference between them. Similarly, most devices related to cars that are simply referred to as being automatic can be described as both *elektrisch* and *elektronisch* e.g. *Fensterheber*, *Wegfahrsperre* (*window mechanism*, *engine immobilizers*). Also a vast number of senses for both adjectives cannot be disambiguated by the definitions provided here. There is *elektronische Kommunikation*, *Anzeigetafeln*, *Daten oder elektronischer Zahlungsverkehr* (*electronic communication*, *display*s, *data or banking*) and which are simply paraphrased as being digital; in other words, electronic in these types of cases basically means 'digital'. With respect to *elektrisch* (*electric*), there are devices and gadgets such as power-driven *elektrische Zahnbürste*, *Nähmaschinen* (*electric toothbrushes, sewing machines*) (most of which are in fact electronic these days too). Real differences occur in collocations, such as *elektrische Aktivität/Impulse* (*electric activity*), which refer to measuring neuronal behavior in the brain, to electric power lines (*elektrische Stromleitung*) being electrically conducted and to physical quantities such as voltage (*elektrische Energie*/*Spannung*).

In essence, the authors tried to capture the main semantic difference (usually one!) between the items in question. As a consequence, readers are given the impression that a word’s meaning is limited to one sense. Polysemous words are not fully explored and where there are in fact a number of differences between them, these are not recorded. As shown in Fig. 9-3, not every description is supported by a citation. However, the examples provided here are always introspective illustrations of contexts. So overall, although its structure is clear and much more user-friendly, most descriptions remain sketchy, with intuitive accounts of meaning.

5. Descriptive, corpus-guided online lexicography: "Paronyme–Dynamisch im Kontrast" (2018)

Despite its established lexicographic tradition covering general monolingual as well as a large number of specialized reference works, there was still no genuine German paronym dictionary in existence in 2018.[11] As lexicographic material and practice underwent a major change in direction towards corpus-based work and electronic editions of dictionaries, a large number of new, more or less reliable resources were developed by publishing houses (e.g. Duden), academic institutions (e.g. IDS Mannheim, Akademien der Wissenschaften Berlin-Brandenburg) and crowdsourcing initiatives (e.g. Wiktionary). However, German linguistics still lacked an empirically sound investigation of paronyms looking at confusables from a native speaker's perspective. The lexicographic landscape also had no adequate dictionary recording the contemporary use of paronyms with editors who retained analytical and editorial control.

A new initiative hosted by the Leibniz-Institut für Deutsche Sprache in Mannheim aimed to close this gap, and a research and documentation project on German paronyms in current language use was undertaken. It focuses primarily on paronyms (*effektiv*/*effizient*, *sportlich*/*sportiv*, *Methode*/*Methodik*/*Methodology*), and provides impressive coverage of word pairs or items of the same word class which are similar in sound/spelling and sometimes in meaning. In 2018, a new, freely accessible online resource called "Paronyme—Dynamisch im Kontrast" was launched, which systematically detects its headwords (cf. Schnörch 2015) and consistently observes regular semantic patterns in language in order to provide comprehensive sense descriptions. This work issues a challenge to existing dictionaries. It represents a radical change from existing German lexicographic traditions by embodying all of the following key features together: a corpus-based foundation, the implementation of meta-lexicographic requirements, a consideration of readers' needs/interests, a concern for the interplay between linguistic and encyclopaedic information and a dynamic, expandable interface with contrastive entries in a linchpin structure.

As a corpus-guided dictionary, it is descriptive in nature, documenting conventionalized patterns and established use without automating the dictionary-making process. It provides information on preferences and

[11] Hausmann (1990, 1123) also points out the absence of such dictionaries: "Reine Paronymwörterbücher sind nach wie vor selten" ("Genuine paronym dictionaries are still rare.").

tendencies as observed in authentic communication rather than following prescriptive traditions. In order to meet meta-lexicographic demands (cf. Rundell 2012; Kövecses and Csábi 2014) a granular XML architecture has been developed so that user-adaptive consultation, dynamic content generation and flexible access routes could be integrated. Users can recreate and re-represent their own dictionary data (cf. Fuertes-Olivera 2013, 330) and it has decisively moved towards a cognitively oriented lexicography (Geeraerts 2007; Ostermann 2015). This means that entries contain linguistic details that are consistently paired up with conceptual-encyclopaedic information. Lexicographic details reflect ideas such as conceptual structure, categorization and knowledge. In this way, the linguistic findings correlate better with how users conceptualize language. Above all, as an online dictionary, space is effectively limitless and hypertextual structures offer new possibilities. In the conceptual phase, consultation modalities were rethought and a flexible approach to information was adopted to meet different needs. As a consequence, users can adapt and generate information dynamically following a number of navigation options, expandable views and menu options. An entry can consist of up to three headwords (e.g. *unsozial*/*asozial*/*antisozial*, *unsocial*/*antisocial*) and it is explicitly designed for contrastive consultation processes. Paronyms are directly compared to each other in visual and explanatory ways. This enables readers to differentiate between definitions, collocates, citations, constructions, sense-related terms and conceptual categories relating to the paronyms as well as the referential domains in which they predominantly occur.

Overall, Hanks (2013, 529) states that there are three aspects which follow the invention of modern computational text processing. These are "compilation, evidence, and use". Here, it is argued that design and presentation have equally been challenged, as these should be effectively in accordance with practical editorial practice. With the help of an example, the characteristics mentioned will be elucidated in this section.

5.1 Corpus data, methods and editorial practice

The empirical corpus-assisted examination of paronym pairs or even sets necessarily incorporates both a suitably large corpus and contrastive meaning analyses. The corpus is a purpose-built collection of data with 2.3 billion words, covering a range of linguistic material between 1995 and 2015.[12] All entries are based on findings from the underlying corpus but this

[12] Following a transparent lexicographic process, this corpus is freely accessible via the corpus management system COSMAS II

does not mean that they are computer-retrieved compilations. In fact, today, pattern-recognition and lexical profiling software supports the systematic structuring of linguistic data with respect to collocations and contexts. As Rundell notes:

> Since the beginnings of corpus-based lexicography, efforts have been made to identify recurrent patterns in the language and to describe them in dictionaries (Rundell 2018, 13).

Manual post-editing of automatically extracted data is necessary to guarantee reliable information. All information is extracted, analyzed and lexicographically interpreted from corpus material. Patterns are interpreted lexicographically; citations and examples are chosen editorially, thus exemplifying authentic conventionalized communication. Usage-based studies and corpus-query systems constantly assist editorial processes.[13]

Besides well-established computational methods, this project also uses a corpus-linguistic method which is capable of measuring semantic similarities (or distance) between pairs of words by contrasting contextual profiles to systematically detect slight differences in terms of collocational behavior. This procedure is referred to as the Contrasting-Near-Synonyms method and was developed by Belica (2001 ff.).[14] It is implemented in a corpus-linguistic collocation research and development workbench (CCDB) (see Belica 2001 ff.; Keibel and Belica 2007) and it offers a visual representation of collocates which resemble the search words to varying degrees in terms of common contexts. On the basis of collocation profiles, semantic structures are analyzed, clustered and visualized in a two-dimensional lattice reflecting different degrees of similarity between various words by using a graded color system. These organizing feature maps (henceforth SOMs) cluster all these items together so that proximity on the grid reflects semantic similarity between semantic profiles.[15] Essentially, this method compares how two words behave semantically by

---

(http://www.ids-mannheim.de/cosmas2/). More information on the Paronym-Corpus can be found here:
http://www1.ids-mannheim.de/lexik/paronymwoerterbuch/dasparonymkorpus.html.

[13] Corpus data is also the foundation of the first systematic headword list of paronyms (cf. Schnörch 2015).

[14] A number of studies of synonymy have successfully employed this method (e.g. Marková 2012).

[15] A similar contrasting method is offered by the feature "Word Sketch differences" within the tool "Sketch Engine" (https://www.sketchengine.co.uk/).

observing all those words that exhibit similar collocation profiles. Vachková and Belica argue that:

> salient SOM features stimulate lexicographers' associative awareness and encourage guided mental imagery leading to valuable insights into both the word semantic structure and the process of discourse-based negotiation of lexical meaning (Vachková and Belica 2009, 239).

As outlined in Storjohann/Schnörch (2014) and in Teichmann (2019) in more detail, feature maps guide lexicographers to those contextual patterns which will provide further evidence, for example, through the study of collocations that can be attributed to specific thematic domains. Generally speaking, the investigation of paronyms often results in a fruitful methodological combination of SOM-based analysis and collocation analyses.

Besides managing the corpus data, tools and corpus-analytic methods, it was also necessary to develop a lexicographic concept with an underlying linguistic theory, as well as a manageable editorial workload. Consequently, this meant leaving the conventional practical concerns of its predecessors aside. Among other things, this dictionary is an attempt to offer reliable and straightforward information as well as intuitive navigation structures. Moreover, its aim is to guarantee a comfortable usability, both for users in their consultation routines and for editors in their daily practical work. The most demanding task was to unite two or three single-lemma descriptions into one homogenous contrastive entry with all its relational elements for comparing and optional (re-)arranging. For this purpose, the writing system incorporates a rather complex XML architecture which is the prerequisite for a dynamic reshuffling of lexicographic information within a two/three-lemma entry. To summarize, it is argued here that it is the interaction between editorial conventions, data requirements and lexicographic expertise that makes a qualified dictionary entry.

5.2 Overview of the entry sportlich/sportiv

Conventionally, inquirers into word meaning could consult existing reference books on confusable terms, provided they were familiar with these, wherein they they will encounter brief normative descriptions of single uses of two confusable lexical items. The central objective of the new resource is to offer contrastive entries of two or three paronym items. Users can instantly familiarize themselves with contextual similarities and differences in concise and clearly arranged comparative entries. Depending on personal interests, the depth of information is realized as a two-level consultation view, i.e. a short overview and an optional detail view.

As Fig. 9-4 demonstrates, different types of context and degrees of similarity are visibly marked using a specific ordering system for contextual uses and a color scheme indicating the type of similarity.[16]

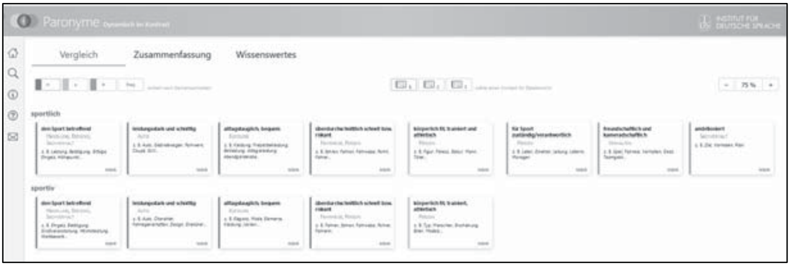


Fig. 9-4: Concise overview entry of similar and different contextual uses of *sportlich*/*sportive.*

Contextual instances are divided into categories based on conceptual differences. The more frequent term occurs at the top of the entry, with all its contextual semantic instances organized in a line. Directly underneath, the second and less frequent, paronym term is listed with its contextual uses. Each 'box' represents a specific contextual use of the adjectives *sportlich* (meaning sporty, sporting, sportsmanlike) and *sportiv* (sporty, athletic). Identical semantic contexts (blue) between the two adjectival items occur to the left of the monitor, followed by similar contexts (green), which, in turn, are followed by different contexts (grey). Instantly, in this entry overview, users can capture semantic parallels, which contexts or discourses they occur in and what these refer to or what they denote. The purpose of this meaning spectrum is to gain an immediate general overview. At this point, each box accommodates a minimum of lexical and encyclopaedic details, a short definition in the form of a synonym, a conceptual reference and prototypical examples (a selection of up to five collocates), e.g.:

- *sportlich* means 'concerning sports' with respect to a PROCESS, EVENT or MATTER, for example *performance*, *pursuit*, *success*, *ambition*, *highlight*,
- *sportlich* means 'powerful and sleek' with respect to a CAR, for example *car*, *Sport Utility Vehicle*, *chassis*, *SUV, coupé*,
- *sportlich* means 'comfortable' with respect to CLOTHES, for example *clothing*, *leisure wear*, *apparel*, *casual clothes*, *evening dress*.

[16] The color scheme will not be visible in a black-and-white print of this article.

This concept-driven navigation structure offers a large amount of knowledge about contextual behavior, parallels or differences, meanings and concepts. Among other criteria, it is the referential/ontological categories underneath the short definitions (e.g. HANDLUNG (PROCESS), EREIGNIS (EVENT), AUTO (CAR), KLEIDUNG (CLOTHES), FAHRSTIL (STYLE OF DRIVING), PERSON (PERSON), VERHALTEN (BEHAVIOR), SACHVERHALT (STATE OF AFFAIRS, SITUATION) that enable users to distinguish patterns and help them to encode/decode contexts and identify metonymic and metaphoric mappings. These are not part of an established ontology or larger onomasiological framework but abstract ad hoc categories. Users can more easily relate the adjectives to their meanings and then relate these to the preferred contextual reference (here modified nouns).

Identical, similar and distinct contexts form separate semantic groups. Users can see which contexts overlap and which meanings are attributed individually. Additionally, the menu options (see buttons at the top left of Fig. 9-4) represent different options for arranging senses. For example, they can be arranged according to overlap, differences or frequencies.[17] In Fig. 9-5, the most frequent senses of *sportlich* and *sportiv* are placed first.

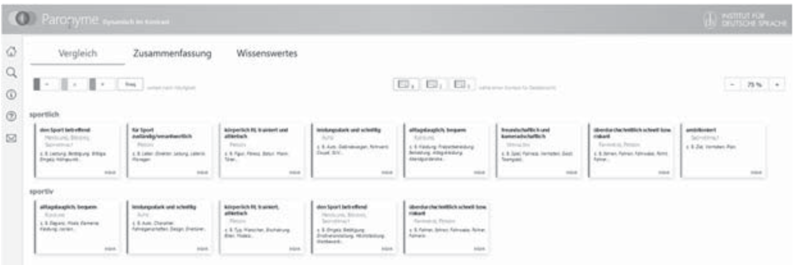


Fig. 9-5: Concise overview entry of contextual uses of *sportlich*/*sportiv* arranged according to frequency.

Accordingly, the positions of the elements change and different facets of structural knowledge are activated on demand.[18] In this way, this reference guide accounts for alternative access routes to language in actual usage events.

---

[17] Sometimes the boundaries between the contextual uses can be blurred, particularly between literal and metonymic/metaphorical uses.

[18] Hanks (2013) points out how difficult it is to decide which meanings to put first in an entry irrespective of how successfully the senses have been distinguished from one another. Re-arranging them according to different paramenters seems a reasonable solution.

5.3 Detail view of sportlich/sportiv

While the overview supports a quick identification of similarities and differences as well as serving as a general navigation guide, the detail view offers more extensive descriptions and citation examples. There is also the option of expanding more data for each contextual use. Users can access more information for a single use of one paronym or, alternatively, up to three contexts can be chosen for contrastive studies. Here, it is the user who decides which ‘box’ to expand and which ‘boxes’ to choose for the parallel consultation of detailed information. An explanatory definition, more typical lexical combinations, relevant topic areas, conceptual categories, corpus examples, synonyms/antonyms and typical constructions in different contexts can be looked up in this more complex detail view. Furthermore, these details can be consulted side by side so that they can be compared easily without switching between information or even entries (see Fig. 9-6).

Fig. 9-6: Contrastive detail view of two contextual uses.

The conceptual references representing encyclopaedic ideas [e.g. EVENT, CAR, PERSON] are explicitly integrated into the more complex definition written in a narrative style. The relevant ontological category (in this context CAR) is specifically illustrated by up to ten collocates, which function as examples of lexical preferences in actual contexts.[19] The details given here are as follows:

[19] In some cases, the conceptual/ontological reference is similar to a domain, as is often the case for nouns and verbs.

sportlich, leistungsstark und schnittig / powerful and sleek
charakterisiert meist ein Auto bzw. dessen Erscheinungsbild dahingehend, dass es z. B. ein tiefergelegtes Fahrwerk sowie stärkere Motorleistung aufweist und optisch schnittig bzw. dynamisch wirkt:
z. B.: *Auto, Geländewagen, Fahrwerk, Coupé, SUV, Limousine, Dreitürer, Flitzer, Optik, Aussehen*.

Engl.: often characterizes a car or its general appearance as being sleek or seeming dynamic, because of its lowered chassis and greater engine output (horsepower)
e.g.: *car, Sport Utility Vehicle, chassis, coupé, SUV, limousine, three-door car, sports car, look, appearance*.

Collocational preferences and typical discursive embeddings help users to decode as well as to encode the word in context. Only then can users successfully and naturally produce language. Users can consult context-dependent linguistic information together with conceptual realizations and discourse-oriented idiosyncrasies. An advantage of this is that it enables them to store linguistic and non-linguistic knowledge together in the mental lexicon. As a result, users can relate the adjective to its meanings more easily and then relate these to the preferred contextual reference (here modified co-occurring nouns). Collocates are treated as family members, as prototypes and fillers in contexts. The information listed in each entry shows a strong interplay between linguistic and extra-linguistic knowledge, and explains how meaning, concept and discourse are intertwined. Taking the example of *sportlich*/*sportiv*, guidewords, conceptual categories and exemplifying domains or reference play an essential role in the identification and navigation of contextual meaning. For lexicographers, these details support the disambiguation process. Patterns of usage encompass conceptual categories as central ideas instead of the sufficient logical criteria of semantic distinction.

The contrastive detail display is not restricted to the comparison of two contextual uses: it is also possible to individually select three contexts. Again, all information assigned to three different senses is listed together in a tabular form to allow for direct comparison. As a matter of fact, such a contrastive detail consultation is not restricted to contexts of different lexical items. It is also possible to individually select two or three contexts for the same word and contrast these in detail. This way, it is possible, for instance, to examine metonymically or metaphorically related contexts of the same lexical item.

5.4 Future lexicographic tasks

It is not only a unit of corpus and cognitive lexicography but also a dynamic resource with flexible and expandable entries. In situations of linguistic doubt, native speakers and learners alike can dip into this resource to resolve their issues. They can learn about thematic domains and semantic environments in which confusable words are likely to occur, together with their natural lexical preferences. Users gain direct access to the meaning and use of two or three easily confused words in concise contrastive overviews or in detailed contrastive descriptions. The entries partly reflect how linguistic knowledge, encyclopaedic knowledge and human experience are inextricably linked in language use and in the mental lexicon. User-adaptable access to details and dynamic search options enable users to gain different perspectives on linguistic information, and authentic examples reflect prototypical structures.

The dictionary currently contains over 360 entries. In the coming years, more entries will be added with editorial care, and specific details will complement existing information. These will include information about semantic change of paronym pairs. Empirical studies of paronyms show how the norms associated with some words have changed over time (cf. Storjohann 2020), including variation in usage between text types and registers (cf. Storjohann 2019), as well as the documentation of irregularities in grammar or oral usage. For that reason, special sections (see under the heading “Wissenswertes”) have been developed, supplementing the existing descriptions of patterns and usage.

6. Conclusion

In this brief survey, I have tried to show that the nature of dictionaries and the paradigms of dictionary-making have changed significantly over time. Corpus linguistics has revolutionized our understanding of how meanings are constructed through context. User studies (e.g. Müller-Spitzer 2014) have increased our focus on users’ interests, and hypertext structures have transformed the presentational style of dictionaries. All these have opened up a whole range of new possibilities in relation to what to include and how to arrange information. Dictionaries have moved away from restricted print editions towards e-dictionaries, and from introspective to empirically sound entries. They no longer regard themselves as an “authority” with an assumed scope of completeness but have, instead, developed into resources of information and knowledge. Nonetheless, “Paronyme – Dynamisch im Kontrast” is unlike both its predecessors and other new dictionaries.

In this new paradigm and with regard to commonly confused terms in German, what used to be a description of a random collection of lemmas has turned into a systematic analysis of potential paronyms. From a lexicographic point of view, we encounter a significant change in conventional practice and the treatment of confusables. Traditional prescriptive approaches have been replaced by new descriptive and more informative accounts. As a corpus-guided dictionary, the primary goal of the new paronym guide is not to account for linguistic norms but to generalize about established usage.

Rundell (2016) claims that a dictionary has three goals:

> From a user's point of view, the ideal dictionary will do three things: it will include the word the user is searching for; it will provide a description of that word's meaning and usage which is accurate and which reflects what the data tells us about its use in real communication; and it will convey this information in a form that takes account of the user's prior knowledge, so that it is both accessible and comprehensible (Rundell 2016, 12).

In terms of the new German dictionary of easily confused words, the following key principles were accounted for. First, comprehensive corpus data and new corpus analysing methods are available to examine lexical pairs in real language use instead of relying on intuitive, instructional approaches paired up with made-up examples and citations. Provided that a corpus is relatively representative and suitable for the lexical tasks, these better reflect language-oriented descriptions which show how paronyms work in authentic communication. In this way, lexicographic descriptions are more reliable sources which foster both linguistic and encyclopaedic information for situations of language doubt.

Second, the depth of information has reached far beyond traditional entries in paper-based reference books. The headword list is a corpus-based account of lemma detection. Roughly 2000 pairs and sets have been systematically identified as potential confusables (cf. Schnörch 2015). In situations of linguistic doubt, native speakers and learners can learn about thematic domains and semantic environments in which confusables are likely to occur, together with their natural lexical preferences. Synonyms and corpus samples illustrate the information provided, while explicit definitions present details which are important for decoding and encoding. Not only is this range of information more objective but it has also surpassed any existing German guide to difficult words.

Finally, as a digital resource, the aim of the new dictionary is to exploit the possibilities of the electronic medium in order to create a flexible, informative and user-friendly instrument which will enable users to make correct choices. This means that instead of rigid, limited and static descriptions, it is now possible to work with dynamic displays and information generated on demand (cf. Kwary 2012, 35). User-specific access and various search options facilitate this resource with the on-demand generated content. The re-positioning of linguistic elements creates different foci and perspectives on the phenomenon and activates different facets of structural knowledge.

Today, users have a free dictionary of easily confused terms at hand where they can expect direct access to the meaning and use of two or three easily confused words together in concise contrastive overviews or in detailed contrastive descriptions. In the future, there will be more opportunities to make dictionaries more appealing to users as long as lexicographers use imaginative approaches and envision new resources. Only then will readers use professional reference works in situations of language doubt instead of googling their linguistic problems.

Bibliography

Adelung, Johann Christoph. 1788. Vollständige Anweisung zur Deutschen Orthographie, nebst einem kleinen Wörterbuch für die Aussprache, Orthographie, Biegung und Ableitung. Leipzig: Weygandsche Buchhandlung.

Belica, Cyril. 1995. Statistische Kollokationsanalyse und -clustering. Korpuslinguistische Analysemethode. Mannheim: Institut für Deutsche Sprache.

Belica, Cyril. 2001ff. Kookkurrenzdatenbank CCDB—V3.3. Eine korpuslinguistische Denk- und Experimentierplattform. Institut für Deutsche Sprache: Mannheim. Accessed at: http://corpora.ids-mannheim.de/ccdb/ [15/12/2019].

COSMAS II: *Corpus Search, Management and Analysis System.* Mannheim: IDS. Accessed at: https://www.ids-mannheim.de/cosmas2/web-app/) [15/12/2019].

Fuertes-Olivera, Pedro A. 2013: "e-Lexicography: The Coninuuing Challenge of Applying New Technology do Dictionar Making", In *The Bloombury Companion to Lexicography*, edited by Howard Jackson, 323-340. London: Bloomsbury.

Geeraerts, Dirk. 2007. "Lexicography." In *The Oxford Handbook of Cognitive Linguistics*, edited by Dirk Geeraerts and Hubert Cuyckens, 1160-1174. Oxford: Oxford University Press.

Gottsched, Johann Christoph. 1762. “Deutsche Sprachkunst: Ein orthographisches Verzeichniß gewisser zweifelhafter Wörter.” In *Johann Christoph Gottsched Ausgewählte Werke*, edited by Phillip Marshall Mitchell (vol. 8.1) and revised by Herbert Penzl [1987], 153-190. Berlin & New York: de Gruyter.

Hanks, Patrick. 2013. “Lexicography from Earliest Times to the Present.” In *The Oxford Handbook of the History of Linguistics*, edited by Keith Allan, 503-536. Oxford: Oxford University Press.

Hausmann, Franz Jospeh. 1990. “Das Wörterbuch der Homonyme, Homophone und Paronyme.” In *Wörterbücher*, vol. 2. edited by Franz Jospeh Hausmann, F.J., Oskar Reichmann and Herbert Ernst Wiegand, 1120-1125. Berlin & New York: de Gruyter.

Keibel, Holger, and Belica, Cyril. 2007. “CCDB. A Corpus-Linguistic Research and Development Workbench.” In *Proceedings of the 4th Corpus Linguistics Conference*, CL 2007, 27-30 July 2007. Birmingham: University of Birmingham, UK.

Klein, Wolf Peter. 2018. Sprachliche Zweifelsfälle im Deutschen. Theorie, Praxis, Geschichte. Berlin & Boston: de Gryuter.

Kövecses, Zoltán, and Csábi, Szilvia. 2014. “Lexicography and cognitive linguistics.” *Revista Española de Lingüística Aplicada,* 27 (1): 18-139.

Kwary, Deny Arnos. 2012. “Adaptive hypermedia and user-oriented data for online dictionaries: A case study on an English dictionary of finance for Indonesian students.” *International Journal of Lexicography,* 25(1): 30-49.

Lăzărescu, Ioan. 1999. Die Paronymie als lexikalisches Phänomen und die Paronomasie als Stilfigur im Deutschen. Bukarest: Anima.

Leibniz, Gottfried Wilhelm. 1983. [1679]. “Unvorgreifliche Gedanken, betreffend die Ausübung und Verbesserung der deutschen Sprache.” In *Unvorgreifliche Gedanken, betreffend die Ausübung und Verbesserung der deutschen Sprache*. *Zwei Aufsätze*, edited by Uwe Pörksen, 5-46. Stuttgart: Reclam.

Marková, Věra. 2012. Synonyme unter dem Mikroskop. Eine korpuslinguistische Studie. Tübingen: Narr.

Mell, Ruth, Schnörch, Ulrich, and Storjohann, Petra. 2019. “Korpussemantische Einflussfaktoren auf Eigenschaften und Funktionen von Paronymen.” *Deutsche Sprache,* 2019 (1): 53-67.

Müller, Wolfgang. 1973. *Leicht verwechselbare Wörter*. Duden Taschenwörterbücher Vol. 17. Mannheim: Bibliographisches Institut.

Müller-Spitzer, Carolin. ed. 2014. *Using Online Dictionaries*. Berlin & Boston: de Gruyter.

Ölinger, Albert. 1573. *Underricht der Hoch Deutschen Spraach*. (Documenta Linguistica 5, Deutsche Grammatiken des 16. Bis 18. Jahrhunderts) (reprint 1975). Hildesheim: Georg Olms.

Ostermann, Carolin. 2015. Cognitive lexicography. A new approach to lexicography making use of cognitive semantics. Berlin & Boston: de Gruyter.

Paronyme—Dynamisch im Kontrast: http://www.owid.de/plus/parowb/. [15/12/2019].

Paronymkorpus: www1.ids-mannheim.de/lexik/paronymwoerterbuch/dasparonym korpus.html. [15/12/2019].

Pörksen, Uwe. 1983. "Der Übergang vom Gelehrtenlatein zur deutschen Wissenschaftssprache. Zur frühen deutschen Fachliteratur und Fachsprache in den naturwissenschaftlichen und mathematischen Fächern (ca. 1500-1800)." *Literaturwissenschaft und Linguistik*, 13. HN 51/52: 227-258.

Pollmann, Christoph, and Wolk, Ulrike. 2010. *Wörterbuch der verwechselten Wörter. 1000 Zweifelsfälle verständlich erklärt.* Stuttgart: Pons.

Rundell, Michael. 2012. "It works in practice but will it work in theory? The uneasy relationship between lexicography and matters theoretical." *Proceedings of the 15th EURALEX International Congress EURALEX 2012*, edited by Ruth Vatvedt Fjeld and Julie Matild Torjusen, 47-92. Oslo: University of Oslo.

Rundell, Michael. 2016. "Dictionaries and crowdsourcing, wikis and user-generated content." In *International Handbook of Modern Lexis and Lexicography,* edited by Patrick Hanks, and Gilles-Maurice de Schryver, 1-16. Berlin: Springer-Verlag.

Rundell, Michael. 2018. "Searching for extended units of meaning—and what to do when you find them." *Lexicography. Journal of Asialex*, March 2018: 1-17.

Sanders, Daniel. 1900. *Wörterbuch der Hauptschwierigkeiten in der deutschen Sprache*. Große Ausgabe. Berlin: Langenscheidtsche Verlags-Buchhandlung.

Schnörch, Ulrich. 2015. "Wie viele Paronympaare gibt es eigentlich? Das Zusammenspiel aus korpuslinguistischen und redaktionellen Verfahren zur Ermittlung einer Paronymstichwortliste." *Sprachreport.* 2015 (4): 16-26.

Sketch Engine: https://www.sketchengine.co.uk/. [15/12/2019].

Storjohann, Petra. 2015. "Was ist der Unterschied zwischen *sensitiv* und *sensibel*?" *Zeitschrift für Angewandte Linguistik,* 62: 99–122.

Storjohann, Petra (ed.). 2019. Paronymie im deutschen Sprachgebrauch. Sonderheft Deutsche Sprache. Berlin: Erich Schmidt Verlag.

Storjohann, Petra. 2019. "Dynamischer Gebrauch von Paronymen in allgemein-, fach-, wissenschafts- und bildungssprachlicher Kommunikation." *Zeitschrift für Angewandte Linguistik*, 71: 1-31.

Storjohann, Petra. 2020. "Paronymie und Sprachwandel." *Muttersprache,* 2020 (2): 144-163.

Storjohann, Petra, and Schnörch, Ulrich. 2014. "Empirical approaches to German Paronyms." In *Proceedings of the XVI EURALEX International Congress: The User in Focus*. 15-19. July 2014, edited by Andrea Abel, Chiara Vettori, and Natascia Ralli, 463-476. Bolzano/Bozen: University of Bozen.

Teichmann, Mareike. 2019. "SOM und CNS als korpuslinguistische Methoden zur Analyse von Paronymen am Beispiel technisch/ technologisch." *Deutsche Sprache,* 2019 (1): 25-41.

Vachková, Marie, and Belica, Cyril. 2009. "Self-Organizing Lexical Feature Maps. Semiotic Interpretation and Possible Application in Lexicography." *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis,* 13 (2): 223-260.

Veit, Werner. 1977. "Rezension zu Wolfgang Müller: Leicht verwechselbare Wörter." *Zeitschrift für Dialektologie und Linguistik*, 44 (2): 225.

Wustmann, Gustav. 1891. Allerhand Sprachdummheiten: kleine deutsche Grammatik des Zweifelhaften, des Falschen und des Häßlichen. Grunow: Leipzig.

CHAPTER TEN

THE MAKING OF THE *HISTORICAL DICTIONARY OF MODERN GREEK*: PROBLEMS AND SOLUTIONS IN THE DOMAIN OF HISTORICAL AND DIALECTAL LEXICOGRAPHY

IO MANOLESSOU AND GEORGIA KATSOUDA

Abstract

The *Historical Dictionary of Modern Greek* (*ILNE*), published by the Academy of Athens, is a long-term multi-volume lexicographic enterprise that aims to document the history of the Modern Greek vocabulary from its origins until today. This paper discusses a) its aims and scope, b) its structure, and c) its methodology, in each case pointing out problems faced during its compilation and how they were solved, in comparison with other Greek and international lexicographic projects.

1. Definition and background

The *Historical Dictionary of Modern Greek, both of the Standard Language and the Dialects* (*Ἱστορικὸν Λεξικὸν τῆς Νέας Ἑλληνικῆς, τῆς τε κοινῶς ὁμιλουμένης καὶ τῶν ἰδιωμάτων* - [istori'kon leksi'kon tis 'neas elini'cis 'tiste ci'nos omilu'menis ce ton iðio'maton]), or *ILNE* for short, is the national lexicographic enterprise of Greece, published by the Academy of Athens (founded in 1927) over a period of several decades. The first volume (vol. 1, *a-am*) appeared in 1933, and since then five more volumes have been published intermittently, with gaps of several years between volumes, caused by consecutive changes in the institution's research and publishing policies. The last volume (vol. 6, *ðe-ðialektos*) was published in 2016, after a major revision and update of the dictionary's corpus and methodology, necessitated by recent progress in historical and dialectal lexicography (fully set out in the new *Manual of Regulations of the Dictionary*, *ILNE-MR*). Currently, the seventh volume is in press (*ILNE*, 2021).

The *ILNE* is primarily a typical historical dictionary, first conceived as a plan in the early twentieth century, by the founder of the discipline of linguistics in Greece, Georgios Chatzidakis[1]. As such, its content and entries include the standard components identified in the typology of historical dictionaries (Merkin 1983; Reichmann 1990a, 1594): a headword, variant forms, part-of-speech and inflectional information, etymology and word history, frequency, collocations, examples from everyday language, quotations from press and literature, synonyms/antonyms and bibliography. Therefore, in many respects it resembles traditional historical dictionaries such as the *New English Dictionary* (now, *OED*), the *Deutsches Wörterbuch* (*DWB*), the *Woordenboek der Nederlandsche Taal* (*WNT*) and the *Svenska Akademiens ordbok* (*SAOB*), which are all long-term multi-volume enterprises, aiming to document a national language in all its detail and in its diachronic dimension. Indeed, the original conceptualization of the *ILNE* was characteristic of late nineteenth century historical lexicography, inspired by nationalist and romantic ideals and aiming to promote a sense of national identity and superiority[2], compounded, in the case of Greek, by the desire to establish a direct connection with the prestigious Ancient Greek civilization. The need was felt as especially pressing, as modern Greece emerged as a modern European state only in the nineteenth century, after 400 years of Ottoman occupation, and as direct contestations of the relation of the modern Greeks to the ancient Hellenes had been expressed by some scholars at the time (especially by the historian Jakob Fallmerayer, cf. Mackridge 2009: 159). On the place of the *ILNE* in the construction of the Modern Greek national identity, see especially Kahane and Kahane (1992, 25-28) and Mackridge (2009, 273-274, 298-299).

---

[1] For the original presentation of the work, see Koukoules (1933), Georgacas and Georgakas (1991), and for the major figure of Georgios Chatzidakis, see Vagiakakos (1977).

[2] On the complex but well-established relationship between historical lexicography and the construction of a national identity, especially in the case of the great lexicographic works of the 19th century, see especially Reichmann (2012: ch. A.5 "Lexikographie in Diensten der Ideologie: Patriotische/nationale, bildungsbürgerliche und volkskundliche Zwecke historischer Wörterbücher"). See also Reichmann (2007, 201-202) and Dollinger (2016).

Despite its categorization as a historical dictionary[3], the *ILNE* differs substantially from other national lexicographical enterprises in one crucial respect: as its title reflects, it covers not only the standard language, but all its dialects as well. This objective multiplies considerably the size of the corpus and consequently lengthens, even more considerably, the time required for the completion of the project. However, it was and remains a necessary decision, rendered inevitable by the special sociolinguistic circumstances of Greek linguistic history.

Modern Greek has a great number of dialects, all (except Tsakonian) deriving from the Hellenistic Koine, which in turn descends from the Attic dialect of Ancient Greek[4]. There is no national dictionary covering all the dialects (i.e. something like the *English Dialect Dictionary* (EDD)) or the *Dictionary of American Regional English* (DARE) for British and American English, respectively. Instead, there are only large dictionaries (synchronic or diachronic) of single dialectal varieties, for example the dictionary of the Pontic dialect (Papadopoulos 1958-1961), of South Italian Greek (Karanastasis 1984-1992) or of Tsakonian (Kostakis 1986-1987)[5]. It is crucial to note that this splintering in the linguistic investigation of the Modern Greek dialects renders historical research difficult, both on a formal and a semantic/lexicographical level, since by definition diachronic linguistics requires comparative data. The issue will be taken up and illustrated later on.

But quite apart from this problem, the main reason behind the necessity of including dialectal data in the historical dictionary of Modern Greek lies in the dual nature of the linguistic history of Greek. Since the first centuries AD, the oral and the written tradition of the language had begun to diverge, eventually leading to a state of diglossia: to oversimplify things, to a contrast between a high-level (conservative, written, official…) and a low-level (vernacular, oral, rapidly evolving…) language variety, which would last for more than 2000 years.[6] In the relevant literature on the topic, Greek is in fact frequently considered as a paradigm case of diglossia

---

[3] See Manolessou (2012) for a fuller presentation of the *ILNE* from the viewpoint of historical lexicography.
[4] For an overview of Modern Greek dialects and their origin, see Trudgill (2003) and Horrocks (2010, 381-412).
[5] A major dictionary of the Cappadocian dialect is also in preparation by the University of Patras in collaboration with researchers from the Academy of Athens, thanks to a state-supported project funded by the Hellenic Foundation for Research and Innovation (HFRI), see http://cappadocian.upatras.gr/en
[6] For overviews of Greek diglossia, see Alexiou (1982), Browning (1982) and Mackridge (1990, 2009). For the influence of the diglossic past in Standard Modern Greek lexicography, see especially Tseronis (2002).

(Ferguson 1959; Hudson 2002). For linguistic research, again, reliance on the high-level variety is not conducive to reliable results concerning diachronic evolution (mechanisms, causes and patterns of change)[7]. But at the time of the *Historical Dictionary*'s inception, the "national language" was just such a high-register, learned, archaic, and to a certain extent artificial, diglossic variety.

This was evident on different levels of analysis: phonology, morphology, syntax and vocabulary. The extremely conservative orthography of Modern Greek (MG) helped to mask the actual phonetic distance between Ancient Greek (AG) and the learned high variety (termed "katharevousa", purist, and contrasting with the "demotic" vernacular), as is evident from the following examples contrasting AG (1a, 2a) with MG purist (1b, 2b) and MG vernacular (1c, 2c) variants:

1a.	ἀνήρ	[an ɛ́:r]	'man'	γυνή	[gynɛ́:]	'woman'	παῖς	[paîs]	'boy'
1b.	ἀνήρ	[a'nir]	'man'	γυνή	[ji'ni]	'woman'	παῖς	[pes]	'boy'
1c.	άντρας	['andras]	'man'	γυναίκαο	[ji'neka]	'woman'	παιδί	[pe'ði]	'boy'

Therefore, for the reliable and in-depth investigation of the history of Greek, recourse to the dialects was inevitable. For example, it frequently happens that a phonetic change appears only sporadically in Standard Modern Greek (i.e. only a small percentage of the vocabulary seems to be affected by it), under the influence of the learned high register language. However, the examination of dialectal data reveals much greater regularity of attestations, as would indeed be expected in the case of a phonetic change. A typical example is the deletion of unstressed initial vowels. The Standard language includes a considerable number of words where the rule applies (2a); but it also includes words which either present variation, with both vowel-initial and consonant-initial forms (2b), and also quite numerous words where the rule does not apply (2c):

2a.	*ὀμμάτιον*	[o'mation]	>	*μάτι*	['mati]	'eye'
	ὀφρύδιον	[o'fryðion]	>	*φρύδι*	['friði]	'eyebrow'

2b.	*ημέρα*	[i'mera]	–	*μέρα*	['mera]	'day'
	ερωτώ	[ero'to]	–	*ρωτώ*	[ro'to]	'I ask'
	υψηλός	[ipsi'los]	–	*ψηλός*	[psi'los]	'tall, high'
	εβδομάδα	[evðo'maða]	–	*βδομάδα*	[vðo'maða]	'week'

[7] On the issue as it affects research on the history of Greek, see Manolessou (2014).

2c.	*εικόνα*	[i'kona]	'image, picture'
	αράδα	[a'raða]	'row, line'
	αρδεύω	[ar'ðevo]	'I irrigate'
	αχλάδι	[a'xlaði]	'pear'
	ελαφρός	[ela'fros]	'light'
	οκνώ	[o'kno]	'I am lazy'

However, if one turns to the dialects, one finds the phenomenon applying with much greater regularity, and affecting all those words which retain the initial vowel in the standard language, see e.g. the following examples from the Cappadocian dictionary under preparation:

3.	*κόνα*	['kona]	'image, picture'
	ράδα	['raða]	'row, line'
	δρεύω	['ðrevo]	'I irrigate'
	χράδι	['xraði]	'pear'
	λαφρό	[la'fro]	'light'
	κνώ	[kno]	'I am lazy'

It is also possible that certain intermediate stages of the linguistic evolution of a word are not attested in the written medieval sources, nor in Standard Modern Greek, but can be found in the oral tradition of the Modern Greek dialects, something which renders them necessary sources for the internal reconstruction of Modern Greek. And in general, the understanding and the study of Medieval, Early Modern and Modern Greek is not possible without the assistance of dialectal data, as was often claimed by the founder of the *ILNE*, Georgios Chatzidakis (Chatzidakis 1905-1907, A 365 and 378).

The inclusion of dialectal material in major historical dictionaries of other languages was always a problematic issue (Reichmann 1990b, 1398-99; Durkin 2015). Although it was understood that dialectal data illuminate language history and constitute evidence both of historical continuity and of the richness and creativity of a language (Wiegand 1998, 690), their inclusion in historical dictionaries was inhibited by mainly ideological reasons. Most historical dictionaries had as one of their aims the promotion of a unified and standardized national language (based on a canon of established authors), something which could not be assisted, but only hindered by the variation exhibited by dialectal data. For Greek, however, this issue did not arise, at least at the time of the creation of the *ILNE*: since the official linguistic standard was the archaizing high register "katharevousa", the inclusion of both common and dialectal forms simply served scientific (comparative-diachronic) purposes, without affecting official language policy. To a certain extent, indeed, it could even be said that this inclusion served the aims of the archaistic policy, since it demonstrated the fragmentation and variability of the spoken language and the absence from it of specialized technical and intellectual vocabulary. This, according to the arguments of the archaists, rendered the spoken everyday language unfit to be used as a standard scientific and administrative tool (a view shared by Chatzidakis himself, cf. the telling title of the paper Chatzidakis 1977 [1926]).

2. Entry structure of the *ILNE* in print and online form

An *ILNE* entry is made up of four parts or building blocks:

(1) The lemma or headword: the headword, outlined in boldface, takes the form of the word in Standard Modern Greek, if it belongs to the standard, or the form of the most conservative variant, if a dialectal word. It may also be an unattested reconstructed form, when all attested variants are innovative and/or widely divergent.

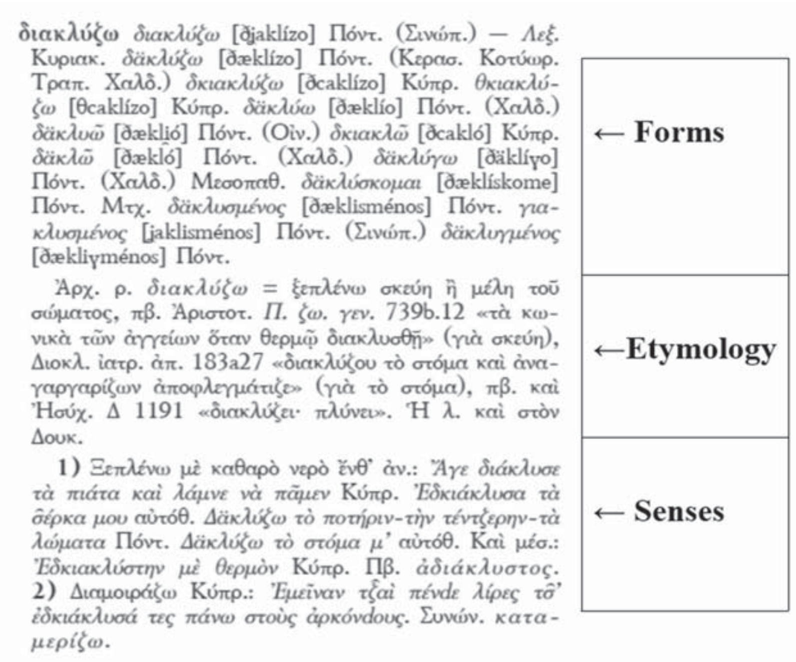
(2) The forms section: this lists all the variant forms of a word in the standard and all the dialects, in chronological order, with phonetic transcription and details of geographical distribution. For new words (nineteenth-twentieth centuries), it also lists the first dictionaries which include the word, as dating indicators for the introduction of the word into the standard language.

(3) The etymological section: this gives the origin of the word, i.e. the ultimate etymon within Greek (or the source language if a loanword), together with the date of first appearance, the original meaning if different from the modern, the dating of variant forms (if no dating is given, it is to be understood that they are first recorded after 1800), and the phonetic and morphological changes that have led to the variants. Although the *ILNE* only includes the vocabulary of Modern Greek (starting from 1800 AD), the history of each word is traced back to its earliest origin within the language. This, given the extremely long history of attestation of Greek (ca. 3000 years) can result in quite lengthy historical accounts, with a multiplicity of diachronic attestations from a variety of sources (ancient and medieval authors, papyri, inscriptions, older lexica and glossaries etc.).[8]

[8] The history of the Greek language can roughly be distinguished in the following periods: Ancient (1500-300 BC), Koine (300 BC-400 AD), Medieval (400-1500 AD), Early Modern (1500-1800 AD) and Modern (1800-). For an overview of periodization, see (*ILNE-MR*, 40). For a synopsis, see e.g. Fortson (2004, 222-244 'ch.12: Greek'), and for a recent full treatment, see Horrocks (2010).

(4) The semantic section: this lists the senses of word in logical (not historical) order, with geographical distribution of each sense, and multiple examples from spoken language, quotations from written sources, as well as synonyms and antonyms. The semantic section also includes observations on register concerning the vocabulary distribution, only within Standard Modern Greek; of course, there can be no question of diastratic differentiation on a dialectal level.

A typical *ILNE* entry can be seen in Fig. 10-1 below:



διακλύζω *διακλύζω* [ðjaklízo] Πόντ. (Σινώπ.) — *Λεξ.* Κυριακ. *δäκλύζω* [ðæklízo] Πόντ. (Κερασ. Κοτύωρ. Τραπ. Χαλδ.) *δκιακλύζω* [ðcaklízo] Κύπρ. *θκιακλύζω* [θcaklízo] Κύπρ. *δäκλύω* [ðæklío] Πόντ. (Χαλδ.) *δäκλυῶ* [ðæklió] Πόντ. (Οἰν.) *δκιακλῶ* [ðcakló] Κύπρ. *δäκλῶ* [ðækló] Πόντ. (Χαλδ.) *δäκλύγω* [ðäklíγo] Πόντ. (Χαλδ.) Μεσοπαθ. *δäκλύσκομαι* [ðæklískome] Πόντ. Μτχ. *δäκλυσμένος* [ðæklisménos] Πόντ. *γιακλυσμένος* [jaklisménos] Πόντ. (Σινώπ.) *δäκλυγμένος* [ðækliγménos] Πόντ.

Ἀρχ. ρ. *διακλύζω* = ξεπλένω σκεύη ἢ μέλη τοῦ σώματος, πβ. Ἀριστοτ. *Π. ζω. γεν.* 739b.12 «τὰ κωνικὰ τῶν ἀγγείων ὅταν θερμῷ διακλυσθῇ» (γιὰ σκεύη), Διοκλ. ἰατρ. ἀπ. 183a27 «διακλύζου τὸ στόμα καὶ ἀναγαργαρίζων ἀποφλεγμάτιζε» (γιὰ τὸ στόμα), πβ. καὶ Ἡσύχ. Δ 1191 «διακλύζει· πλύνει». Ἡ λ. καὶ στὸν Δουκ.

1) Ξεπλένω μὲ καθαρὸ νερὸ ἔνθ' ἀν.: *Ἄγε διάκλυσε τὰ πιάτα καὶ λάμνε νὰ πᾶμεν* Κύπρ. *Ἐδκιάκλυσα τὰ σ̂έρκα μου* αὐτόθ. *Δäκλύζω τὸ ποτήριν-τὴν τέντζερην-τὰ λώματα* Πόντ. *Δäκλύζω τὸ στόμα μ'* αὐτόθ. Καὶ μέσ.: *Ἐδκιακλύστην μὲ θερμὸν* Κύπρ. Πβ. *ἀδιάκλυστος*. **2)** Διαμοιράζω Κύπρ.: *Ἐμεῖναν τζ̂αὶ πένdε λίρες τσ̂' ἐδκιάκλυσά τες πάνω στοὺς ἀρκόνdους*. Συνών. *καταμερίζω*.

Fig. 10-1: *ILNE* entry for the verb *διακλύζω* [ðʝa'klizo] 'to rinse', first attested in Aristotle, absent from Standard MG, and preserved only in the dialects of Cyprus and Pontus.

Of course, words which belong to the basic vocabulary, both of the standard and of the dialects, are treated in much larger entries, where each section (except for the headword) may take up several pages of printed text, due to the long history and extensive regional variation of the word. Fig. 10-2 below gives, as an example, just part of the entry *δέρνω* ['ðerno] 'to beat up' < AG *δέρω* [dérɔ:]:

δέρνω *δέρω* [ðéro] Ἀγαθον. Αἰτωλοακαρν. (Φτελ.) Ἤπ. ἐ.Ἑλλ. (Ἅγιοι Σαρ. Δέλβ. Δερβιτσ. Χιμάρ.) Θεσπρωτ. (Κεστρ. κ.ἀ.) Ἰωανν. (Βούρμπ. Πυρσ. Πωγών. κ.ἀ.) Καρ. (Ἁλικαρν.) Κάρπ. Κύπρ. (καὶ *δέρνω*) Νίσ. (καὶ *δέρνω*) Πάρ. Πόντ. (Ἀμισ. Σάντ. Σταυρ. Τραπ. κ.ἀ.) Σύμ. Τσακων. (Χαβουτσ.) Φοῦρν. κ.ἀ. — περιοδ. *Πανδώρα* 9 (1858), σ. 215 περιοδ. *Ν. Ἑστ.* 210 (1935), σ. 850 *δέρου* [ðéru] Ἄρτ. (Κομπότ. κ.ἀ.) Γρεβεν. (Γρεβεν. κ.ἀ.) Ἕβρ. (Διδυμότ.) Θεσσ. Θράκ. ἐ.Ἑλλ. (Ἀβδήμ. Ἀδριανούπ.) Ἰωανν. (Ζαγόρ. Ἰωάνν. Κουκκ. Λάκκα Σουλ. Ξηροβ. κ.ἀ.) Πρεβ. (Γοργόμ. Πρέβ. κ.ἀ.) Τρικ. (Μυρόφ.) Τσακων. (Πέρα Μέλαν. Πραστ. Τυρ.) — Abbott (1900: 123) *ἔρω* [éro] Κάρπ. Κάσ. Πάτμ. Χάλκ. *γέρω* [jéro] Κάρπ. *δείρω* [ðíro] Ἀχαΐας (Αἰγιαλ.) Κύπρ. *δείρου* [ðíru] Τσακων. *δέρνω* [ðérno] κοιν. καὶ Καππ. (Σινασσ.) Πόντ. (Τρίπ. κ.ἀ.) *δέρνου* [ðérnu] βόρ. ἰδιώμ. Λυκ. (Λιβύσσ. Μάκρ.) *ἔρνω* [érno] Κάλυμν. Κάρπ. Κῶς (Ἀντιμάχ. Κέφαλ. κ.ἀ.) Ρόδ. (καὶ *δέρνω*) *γέρνω* [jérno] Κάρπ. Κύπρ. *ζέρνω* [zérno] Κάλυμν. Ἀόρ. *ἔδειρα* [éðira] κοιν. *ἔδερα* [éðera] Κύπρ. (Λευκωσ. Πεδουλ.) Μεγίστ. Νίσ. Πάρ. Ρόδ. Σκύρ. — Παλαίσ. *Συλλ. ποιημ.*, σ. 151 *ἔγερα* [éjera] Κύπρ. (Κοίλ. κ.ἀ.) Ρόδ. *ἤερα* [íera] Κάρπ. *ἔδαρα* [éðara] Ἀχαΐας (Αἰγιαλ.) *ἐδάρκα* [eðárka] Τσακων. *δάρκα* [ðárka] Τσακων. (Χαβουτσ.) Προστ. ἀορ. *δεῖρες* [ðíres] Ἰωανν. (Βούρμπ.) Κοζ. (Κοζ.) Μτχ. *δερνῶdας* [ðernódas] Σκύρ. Μεσοπαθ. *δέρομαι* [ðérome] Ἤπ. ἐ.Ἑλλ. (Ἅγιοι Σαρ. Δέλβ. Σχωρ.) Κάρπ. Λευψ. Πάτμ. Σύμ. *δέρουμι* [ðérumi] Ἰωανν. (Δολιαν.) *δέρκουμαι* [ðérkume] Πόντ. (Ὄφ. Σταυρ. Τραπ. κ.ἀ.) *ἔρομαι* [érome] Κάρπ. *δερούμενε* [ðerúmene] Τσακων. (Πέρα Μέλαν. Πραστ. Τυρ. κ.ἀ.) *δέρνομαι* [ðérnome] κοιν. καὶ Καππ. (Σινασσ.) *δέρνουμαι* [ðérnume] Ἀρκ. (Βούρβ. Δημητσ.) Ζάκ. Ἠλ. (Φιγάλ.) Θράκ. ἐ.Ἑλλ. (Σαρεκκλ.) Κέρκ. Μεσσην. (Μανιάκ.) Πόντ. (Τραπ.) Ρόδ. *δερνοῦμαι* [ðernúme] Πάρ. *δέρνουμι* [ðérnumi] πολλαχ. βορ. ἰδιωμ. *δέρνομου* [ðérnomu] Παξ. *δέρνουμου* [ðérnumu] Λυκ. (Λιβύσσ. Μάκρ.) *δερνίομαι* [ðerníome] Πόντ. *δερνίουμαι* [ðerníume] Πόντ. *διρνίομαι* [ðirníome] Πόντ. (Οἰν.) *δάρκουμαι* [ðárkume] Πόντ. (Ἀμισ. Ἴμερ. Ὄφ. Τραπ. Χαλδ. κ.ἀ.) *δάρχκουμαι* [ðárxkume] Πόντ. *δάρχουμαι* [ðárxume] Πόντ. *δαρκούμενε* [ðarkʰúmene] Τσακων. (Πέρα Μέλαν. Πραστ. Τυρ.) Πρτ. *ἐδέρκουμουν* [eðérkumun] Πόντ. (Τραπ.) Ἀόρ. *δάρθηκα* [ðárθika] κοιν. *ἐδάρου* [eðáru] Κύπρ. *ἐρδάα* [erðáa] Πόντ. *ἐδάρ-*

Μεσν. ρ. *δέρνω*, τὸ ὀπ. ἀπὸ τὸ ἀρχ. ρ. *δέρω*, ὅπου καὶ τύπ. *δείρω* = α) ἀφαιρῶ τὸ δέρμα (σφαγμένου ζώου), γδέρνω β) κτυπῶ, μαστιγώνω, κατὰ ἀναλογ. μεταπλ. τῶν ὑγρολήκτων ρ. πρὸς ἄλλα σύγχρ. σὲ *-νω*. Βλ. Χατζιδάκις, ΜΝΕ Α΄, 291-292, Horrocks (2010²: 305-306), καὶ πβ. *φέρω-φέρνω*. Ἡ λ. καὶ στοὺς Μεούρσ. Βλάχ. Δουκ. (λ. *δέρνειν*) Σομ. Οἱ διαλεκτικοὶ τύπ. χωρὶς τὸ *-ν-* ἀπευθείας ἀπὸ τὸ ἀρχ. *δέρω*, ἀλλὰ ὁ τύπ. *δέρου* σὲ πολλὰ βόρ. ἰδιώμ. πιθ. ἀπὸ νεότ. ἁπλοποίηση τοῦ [n] σὲ περιπτώσεις δημιουργίας πολυμελῶν συμφωνικῶν συμπλεγμάτων, ὅπως *δέρν'ς, δέρν'ν*, βλ. Ἀναγνωστόπουλος (1924: 74-75). Γιὰ τὴν τροπὴ [ð] > [z] στὸν τύπ. *ζέρνω* βλ. λ. δ 4β καὶ Παντελίδης (1929: 38). Ὁ τύπ. *ἔδερα* ἤδη μεσν. (πβ. *Ἀχιλλ.* L 410 «ἀλλάγια δέκα ἔδεραν καὶ ἐξέβησαν ὡς ἄνδρες») ἀπὸ τὸν τύπ. *ἔδειρα* μὲ κανονικὴ τροπὴ [i] > [e] λόγῳ τοῦ γειτνιάζοντος ὑγροῦ [r]. Ὁ τύπ. *ἔδαρα* ἀπὸ τὸ θ. τοῦ μεταγν. ἀορ. *ἔδαρον* (Ἡρωδιαν. *Π. ῥημ.* 3.2.801 «δείρω ἔδαρον ἐξ οὗ τὸ ἐδάρην, τέμνω ἔταμον»). Ὁ ἀόρ. *ἐδάρκα* μὲ τὸ ἐπίθμ. τοῦ ἀρχ. πρκ. *-κα* ἢ καὶ κατὰ τοὺς ἀορ. σὲ *-κα*, ὅπως *ἔδωκα, ἀφῆκα* κ.τ.τ., πβ. Κωστάκης (1951: 99-100, 181-182). Ὁ τύπ. *δερνῶdας* ἀπὸ ἀναλογία πρὸς τὶς μτχ. ἐνεστ. τῶν συνηρ. ρ. (π.χ. *περπατῶντας*) καὶ τὶς μτχ. ἀορ. β΄ (π.χ. *δραμόντας, διαβόντας*)· βλ. Χατζιδάκις, ΜΝΕ Β΄, 122-123 καὶ πβ. *γυριζῶντας* < *γυρίζω*, *ἐρχῶντας* < *ἔρχομαι*. Ὁ τύπ. *δέρκουμαι* μὲ τὸ ἐπίθμ. *-κουμαι* < *-γουμαι*, γιὰ τὸ ὀπ. βλ. Οἰκονομίδης (1958: 315-317). Ὁ τύπ. *δέρνομαι* καὶ στοὺς Βλάχ. Σομ. Ὁ τύπ. *διρνίομαι* μὲ στένωση τοῦ [e] σὲ [i], γιὰ τὴν ὀπ. βλ. Οἰκονομίδης (1958: 24). Οἱ τύπ. *δάρκουμαι, δάρχκουμαι, δάρχουμαι* ἀπὸ τὸ θ. τοῦ παθ. ἀορ. *ἐ-δάρθ-α*. Γιὰ τὸν σχηματ. τοῦ τύπ. *ἐδάρμα* βλ. Anagnostopulos (1926: 62), Κωστάκης (1951: 107). Ὁ τύπ. *ἐρδάα* ἀπὸ τὸ ἀρχ. *ἐδάρην*, ὅπως καὶ τὸ *ἐχάρην-ἐχάρα* διὰ τῶν μτβ. τύπ. **ἐδάρα* > **ἐδράα* > *ἐρδάα* μὲ μετάθ. τοῦ [r]. Ὁ τύπ. *ἐνdράρθα* ἀπὸ ἀμάρτ. *ἐνdάρθα* μὲ ἀφομοιωτ. ἀνάπτ. [r], πβ. *βουκέντρι-βουρκέντρι*. Πβ. καὶ τύπ. *ἐδάρθηκα* καὶ στὸν Σομ. (λ. *δέρνομαι*). Γιὰ τὸν τύπ. *ἐδέρτησα* < *ἐ-δέρ-θησα* βλ. Μηνᾶς (2004: 723-724). Ὁ τύπ. *δαρθῶ* καὶ στὰ Σχόλ. Θεόκρ. 18.9b «κατέδαρθες: τὸ θέμα καὶ δάρθω

Fig. 10-2: Part of the *ILNE* entry for *δέρνω* ['ðerno] (vol. 6, p. 160).

The *ILNE*, given its early beginnings, is a printed dictionary, but in the last years the Academy of Athens has implemented a project of retro-digitization of its older volumes, in machine-readable form. Retro-digitization has faced a number of problems, mainly due to the special custom-made symbols and diacritics used in the dictionary for the representation of sounds not extant in the standard language, which are not easily recognizable by OCR programs (Fig. 10-3).

α̈ = [æ]	ḍ = [d]
σ̂ = [ʃ]	τ̔ = [tʰ]
κ́ = [tɕ]	ζ̆ = [ɾ̞]

Fig. 10-3: (Some of the) special symbols used in the *ILNE.*

The problem was compounded by the fact that the meta-language of the earlier volumes (until the new Regulations in 2012; see *ILNE-MR*, p. 3) was not Standard Modern Greek, but an archaizing, high-register form of katharevousa, also not included in standard OCR packages as an identifiable language. The problem was resolved through a combination of Ancient Greek and Modern Greek OCR, followed by post-correction by a large team, through outsourcing. Newer lexicographical enterprises involving the Academy of Athens (such as the Cappadocian dialectal dictionary mentioned above) naturally do not face such problems any longer, as they are implemented on a digital basis from the start, using exclusively Unicode fonts and state-of-the-art DWS (dictionary writing systems) tools, in order to facilitate, streamline and speed up the writing process, and the meta-language is of course Standard Modern Greek (see now Karasimos, Manolessou & Melissaropoulou 2020).

Thanks to the retro-digitization enterprise, volumes 1 through 5 of the *ILNE* are now available online, at the following link: http://repository.academyofathens.gr/kendi/index.php/en, as part of the Digital Depository of the Academy of Athens. The interface, through the "Advanced Search" button (Fig. 10-4a), offers the user the opportunity to search either for a dictionary entry via a headword search, or for a specific local variant (through a search of the forms section), or for a specific area (e.g. for all the words contained in the dictionary from a specific settlement, island, or administrative unit) or for a specific meaning. The user can also download the page in question, or even the whole volume, as a .pdf file (Fig. 10-4b). The most recent volumes, volume 6 (published 2016) and volume 7 (2021) are planned to circulate only in print form for a short number of years, before being uploaded into the digital database.

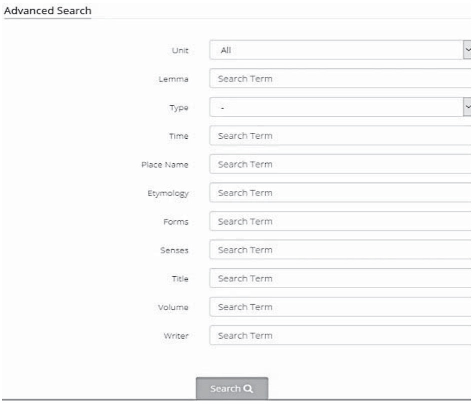


Fig. 10-4a: Advanced Search Interface of the *ILNE*.

Fig. 10-4b: Search Results: download page or read/ browse online.

3. Phonetic transcription

The synchronic examination of dialectal data in the *ILNE* involves a set of problems having to do both with the reliability dialectal data themselves and their representation. A basic aspect of these problems is phonetic transcription. The phonological subsystems of the various MG dialects include sounds that do not exist in Standard MG, and therefore cannot be represented with the Greek alphabet and its spelling conventions. Such sounds are: double (geminate) consonants (6a), various types of palatalization (6b) and various types of laterals (6c), as illustrated below in Fig. 10-5:

(a) *ποτέ* ‘never’	(b) *και* ‘and’	(c) *καλά* ‘well’
ποτέ [poˈte] Standard MG	και [ce] Standard MG	καλά [kaˈla] Standard MG
ποτ-τ́έ [poˈtʰ:e] Cyprus	τσ̑αι [tʃe] Cyprus	καƛά [kaˈʟa] Macedonia
ποτθέ [poˈtᶿe] Kalymnos	κ́αι [tɕe] Rhodes	καλ̣ά [kaˈɻa] West Crete

Fig. 10-5: Realization of three words in Standard MG and in dialects.

The problem has two aspects:

(1) How to represent dialectal sounds in the *ILNE*: the exclusive use of the International Phonetic Alphabet would render the dictionary virtually unreadable for non-expert linguists, not only in the forms section, where the variant regional forms are listed, but especially in the examples and quotations section. It must be stressed that the *ILNE*, as a national Historical Dictionary, is not addressed only to the linguistic community, but also to students, philologists, historians and, to a certain extent, the general public, i.e. to readers without special linguistic training[9]. To that effect, it has developed a custom system of notation, using letters of the Greek alphabet with special diacritics. This practice is well-established in all European countries, as the International Phonetic Alphabet does not have a long tradition in the field of dialectology and dialectal lexicography, where various alternative transcription systems, such as the Teuthonista or the Rousselot-Guilléron alphabets, are dominant[10]. Nevertheless, as stated in the new Regulations of the *ILNE* (*ILNE*-MR), it has been deemed necessary to also include, in the formal section, a standard IPA transcription for all variant forms (as can be seen in Fig. 10-1 and Fig. 10-2 above). The gain from such a measure is obvious: The transcription will be transparent, consistent and standardized, and the data will be accessible to the international scholarly public, not only to Greek-speaking scholars.

(2) How to use the IPA when transcribing written dialectal sources. Much of the primary material of the dictionary comes from fieldwork conducted in past decades, without sound recordings, and from older dialectological studies which do not employ the IPA. These often employ an *ad hoc* symbolism which is no longer comprehensible, or describe phonetic phenomena with non-standard, antiquated or amateur terminology (such as ‘cacuminal’, ‘heavy’, ‘thick’, ‘muted’). When transcribing a word,

---

[9] On the educational role of historical lexicography in society and the concomitant necessity of accessibility to a wide range of disciplines, see Reichmann (2012: ch. A.4).

[10] For an overview of dialectological practice with respect to phonetic transcription, see Manolessou, Beis and Bassea-Bezantakou (2012), and Schiefer and Pompino-Marschall (1996).

the *ILNE* researcher is thus frequently forced to simply guess what the older source must have meant. In principle, one could of course decide not to include in the dictionary any data not corroborated by modern phonetic research and backed up by sound recordings. But this step would entail the rejection of about 80% of the collected material, which in most cases is unique or rare and hard to find, and comes from dialects nowadays close to extinction. Furthermore, even if this material is unreliable phonetically, it still provides invaluable morphological and semantic information which cannot be ignored in a historical dictionary. The *ILNE* cannot provide bibliographic references for each dialectal form listed (for reasons of space, as this would increase the size of the dictionary more than a hundredfold) so that the interested scholar could evaluate the reliability of the transcription of the form; however, since it follows the principle of information recoverability, the provenance of all forms could potentially be traced back by visiting the *ILNE* archive, where the examples are excerpted, previously on card-slips and recently in digital form.

In order to transcribe the bewildering variety of amateur phonetic symbols occurring in the older sources as consistently as possible, one of the preparatory tasks of the *ILNE* was to compile correspondence tables for the symbolism used by the major dialectological studies for each dialect, and to set down strict rules for the use of the custom-made extra symbols complementing conventional orthography. The symbols in each study were correlated to each other, and given a correspondence in the custom notation of the *ILNE* and in the IPA. To give a concrete example, Fig. 10-6 below reproduces part of the comparative table for the Pontic dialect:

Παρχαρίδης 1885	Βαλαβάνης 1892	Οικονομίδης 1908	Παπαδόπουλος 1955, 1958	Οικονομίδης 1958	Συμεωνίδης-Τομπαΐδης 1999	ILNE	IPA
ʿα	ια	ᾰ̈	α̤	ᾰ̈	α̤	ᾰ̈	æ
ʿο, ʿω	ιο, ιω	ŏ̈, ω̈	ο̤	ŏ̈, ω̈	ο̤	ŏ̈, ω̈	ø
ʿου	ιου	ŭ̈		ŭ̈, ü		ŭ̈	y
ᵒι, ᵒε		ŭ		ı		i	ɯ
h	h	h	h	h	h	h	h
	p, t, k		p, t, k	π̇, ṫ, κ̇, κ̈		π̇, ṫ, κ̇,	pʰ, tʰ, kʰ, cʰ
		κ̂		κ́	κ́	κ̂	c
μπ	b	β̇, π̇		b	b	μπ, b	b
ντ	d	ṫ, δ̇		d	d	ντ, d	d
γκ	g	κ̇		g	g	γγ, γκ, νg, g	g
		κ̇̂		ǵ	ǵ	ĝ	ɟ
		χ̂		χ́		χ̂	ç
		γ̂		γ́		γ̂	j
j	j	j, γ̂		j	j	γ, j	j

Fig. 10-6: Alternative phonetic notations for the Pontic dialect.

4. Historical and etymological investigation in the *ILNE*

The *ILNE* attempts the full investigation of the Modern Greek vocabulary on two axes, diachrony and synchrony. The first has to do with the etymology and historical evolution of each word, while the second has to do with its regional variation but also with its use in Standard Modern Greek (everyday speech, text corpora, literature, the press).

The historical and etymological treatment of a word takes place mainly in the etymological section, although some historical information may be found in the other two sections as well (chronological-evolutionary ordering of forms in the formal section, and dating of senses in the semantic section). The etymological principles of the *ILNE* are set out in articles 79-94 of the *New Manual* (*ILNE- MR*, see also Manolessou 2012). Its basic etymological principles are the following (exemplified through data taken from volume 6 (2016) and volume 7 (2021):

4.1. Research on the basis of primary sources

The *ILNE* relies on the exhaustive research of primary sources, and avoids second-hand lexicographical information. This frequently leads to revisions to the historical and etymological information provided in other dictionaries. A characteristic example from volume 6 is the word *δευτερόλεπτο* [ðefteˈrolepto] 'second, unit of measure of time'. The ancient meaning is 'second, unit of measure in astronomy or geometry'. Standard MG dictionaries date the semantic shift of the word to the modern period, and attribute it to influence from French; however, primary research in digital historical corpora of the Greek language (mainly the TLG—*Thesaurus Linguae Graecae*), shows that the new meaning has appeared in medieval times, as is evident from the following quotation, a scholium on Hesiodus by the Byzantine scholar Ioannes Tzetzes (4):

4. «λέγω δὲ, καιροὺς, μῆνας, ἑβδομάδας, ἡμερόνυκτα, ὥρας, λεπτὰ, δευτερόλεπτα» *Scholia Tzetzae in Hesiodi Opera et dies,* v. 613 (Gaisford 1823)
   "I mean chronological periods, months, weeks, days and nights, hours, minutes, seconds".

Similarly, the word δικέφαλος [ðiˈcefalos] 'two-headed' has a specialized sense as a medical term, namely 'biceps'. This sense is interpreted as a semantic calque from French in several Modern Greek dictionaries, but can in fact already be found in Byzantine medical treatises (5):

5. «ἡ ἀγκυροειδὴς τῆς ὠμοπλάτου ἀπόφυσις· ᾗ ἐπίκειται μῦς ὁ λεγόμενος δικέφαλος» Meletius, *De Natura hominis,* 119.1 (Cramer 1836)
"The anchor-like edge of the shoulder, to which is attached the muscle called biceps"

4.2 Checking of all dictionaries—detailed treatment of loanwords

The process of loanword treatment can be exemplified through two opposing examples. The first is an apparently Greek dialectal word, *διανώρα* [ðjaˈnora] 'continuously' (attested in the island dialects of Naxos, Syros, Cythera and Chios), which entered into the first volume of the *ILNE* in the form *ἀδιαώρα* [aðjaˈora], on the basis of a quite probable internal Greek etymology, the prepositional prefix δια- [ðia] 'through' and the noun *ὥρα* [ˈora] 'hour' < AG *ὥρα* [hɔ́:ra] 'season, time of day', with prothetic [a]. The etymology was reinforced through comparisons with other compounds consisting of a prepositional prefix and the noun ώρα [ˈora], such as *ἐξ-ώρας* [eˈksoras] 'late', *προς-ώρας* [proˈsoras] 'temporarily'. However, more extensive recent lexicographic work for the new volume of the *ILNE* has shown that it is in fact a loanword from Venetian, and specifically from the phrase *da ogni hora* 'all the time' (cf. Cortelazzo 2007, s.v. *dagnóra*). Its attestation only in areas previously (thirteenth-sixteenth century) under Venetian occupation was another indication of its loanword status.

The reverse example is of special interest for the history of linguistics, as it involves the origin of a well-known linguistic term, namely *diglossia*. The word *διγλωσσία* [ðiɣloˈsia] 'diglossia' is first attested in Greek Christian texts of the first century AD, with the meaning "deceitfulness". This is the meaning it retains throughout the Medieval and Early Modern period. But when does the modern linguistic meaning "a situation where two very different varieties of a language co-occur throughout a speech community, each with a distinct range of social function"[11] appear? All dictionaries of Standard Modern Greek agree that this is a recent loan translation from French *diglossie*, basing this claim on Ferguson's classic article "Diglossia", where in fact he states that he borrowed the term from French linguistics, because no English equivalent exists: "The term 'diglossia' is introduced here, modeled on the French *diglossie*, which has been applied to this situation, since there seems to be no word in regular use for this in English" (Ferguson 1959, 325-326). However, a more careful perusal of Ferguson's references, and a more detailed research on the basis of the relevant linguistic literature, reveals that the earliest French works

[11] The definition is from Crystal (2003, s.v.).

that Ferguson drew on in fact described Modern Greek (given that, as discussed above, Greek constitutes a paradigm case of diglossia), and were written by French scholars (e.g. Pernot 1897) adopting and repeating an innovative use of the word first appearing in the Greek language. Therefore, the first appearance of "diglossia" with the new linguistic meaning can be traced back to the Greek *διγλωσσία*, as used by the scholars Yannis Psichari and Emmanouil Roidis—the earliest attestation with this meaning dates back to 1885[12].

Entry	Koumanoudis (1900)	*ILNE*
δημοκράτης [ðimo'kratis] 'democrat'	1839	1797
δημοπρασία [ðimopra'sia] 'auction'	1840	1807
διαιώνισις [ðie'onisis] 'perpetuation'	1862	1818
διεύθυνση [ði'efθinsi] 'direction'	1833	1786
διευκολύνω [ðiefko'lino] 'to facilitate'	1857	1814
δίζυγο ['ðiziɣo] 'parallel bars'	1896	1837
Διμηνία [ðimi'nia] 'period of two months'	1871	1824

Table 10-1: Ante-dating in the *ILNE*.

The combination of these two research principles has led to a consistent re-dating of many words in the Modern Greek vocabulary, for which the *ILNE* provides quite different chronological data than all other dictionaries of Modern Greek. This is especially true of the so-called "higher-register" or "learned" vocabulary, which is often erroneously dated too late, usually on the basis of a single lexicographical work of the beginning of the last century, Koumanoudis (1900), which provided a list

[12] The issue is discussed in detail in Fernández (1995) and Mackridge (2007, 27-31).

of the first appearance of neologisms (termed in the title of the work as "new words"). Table 10-1 lists a number of high-register words which are consistently post-dated in Standard Modern Greek dictionaries on the basis of the data provided by Koumanoudis, and which are assigned an earlier date through the research conducted for the *ILNE*, volumes 6 and 7. The case of *δίζυγο* [ˈðiziɣo] 'parallel bars' is instructive: Koumanoudis (1900) dates it to 1896, the date of the revival of the Olympic Games (the first Olympic Games of the modern era, which took place in 1896 in Athens). In fact it occurs in gymnastics manuals addressed to teachers in Greek state schools as early as 1837 (the date of the foundation of the new Greek state and of state education).

4.3 Discussion of alternative etymologies

A third basic principle adopted by the *ILNE* is the full mention of alternative etymological proposals, with bibliographic references. Full referencing is usually dispensed with in the etymological section of major monolingual dictionaries of the standard language, as a concession to brevity (cf. Drysdale 1989). Of course, a historical dictionary cannot provide as full an analysis as an etymological dictionary (for the difference between the two, see e.g. Buchi 2016), but in the recent volumes there is a consistent effort to provide full bibliographic data in order to direct the reader to further relevant research. Intriguing individual cases which cannot be treated adequately within the narrow framework of a dictionary entry are occasionally discussed in detail in one of the periodicals issued on an annual basis by the Academy of Athens, namely the *Λεξικογραφικόν Δελτίον* [leksikoɣrafiˈkon ðelˈtion] 'Lexicographical Bulletin'.

4.4 Documentation of dating and etymological proposals through quotations

The *ILNE* is the only Modern Greek dictionary that documents its definitions and datings through authentic corpus examples, both of oral speech (from recordings and transcriptions of dialectal fieldwork, from textual corpora of the standard language such as the HNC and from the Internet - "web as corpus"), and from written language (digital depositories of newspapers and periodicals, works of Modern Greek literature, checked from reliable editions).[13]

[13] The sources of the *ILNE* are discussed briefly in *ILNE*-MR and in detail in Konstantinidou and Tzamali (2012).

Quotations from literary works are almost entirely absent from other major dictionaries of Modern Greek, as a result of the biased view that literary quotations represent artificial or "deviant" language. But for a historical dictionary, quotation evidence constitutes one of its major assets, the "bedrock" and the "pre-requisite" of historical lexicography in the terms of Hawke (2016, 176 and 202). This is for a variety of reasons: a) for the secure dating or for the wide range of illustration it can provide, b) as a documentation of word history, since major figures of literature frequently set the trend for the adoption or spread of lexical innovation, and c) as learning aids, since a new word or meaning can more easily be imprinted in the reader's mind through a striking, memorable or well-known literary excerpt. And, of course, a historical dictionary also needs to function as a depository and anthology of the language's major literary production, and lead the reader to the discovery of high-quality national literature, both of the past and of the present (on this issue, see Considine 1998)[14].

4.5. Comparative examination of dialects

As a fundamental principle of the *ILNE*, accurate dating and successful etymology can only be achieved through comparative examination of the totality of available dialectal data (consisting of older collections of material, recorded and written, as well as older and more recent local dialectal glossaries and dictionaries, both scholarly and amateur—see fn. 13). This is an issue that was briefly touched upon earlier, and which requires further discussion and illustration. The long history of the Greek language and the lack of documentation for evolutions in the vernacular spoken language (due to the phenomenon of diglossia described above) may result in isolated dialectal forms whose provenance, etymology and gradual development are hard to determine. In such cases, the answer may be found by examining the data of a widely distanced dialect (both in spatial and in temporal terms). To give two concrete examples: the word *δίδυμος* ['ðiðimos] 'twin' has lost its original meaning in the Greek dialect of Calabria, in South Italy. It now only means "an error in the weaving of cloth, whereby two threads pass through the same hole". The *Historical Dictionary of South Italian Greek* asserts that this new sense is unattested in other Greek dialects (Karanastasis 1984-1992, s.v. *δίδυμο*), and thus its dating cannot be determined; it could be considered a recent South Italian semantic innovation. However, comparative research for the *ILNE* has shown that the new sense in fact appears also in two other areas, the island of Skiathos and

[14] Literature quotations in the *ILNE* are discussed in Katsouda (2016).

in Pontus, Asia Minor, which, crucially, have not had any contact with Calabria since the eleventh century (the time when this territory of the Byzantine Empire was conquered by the Normans). Therefore, the new research can both confirm the meaning of the Calabrian word, and determine a dating to the Middle Ages, despite the lack of written documentation. In the same vein, the word *διβόλι* [ði'voli] is a very common agricultural term, attested in dozens of dialects, and usually has the meaning 'ploughing of a field for the second time, double ploughing'. However, there are only two areas where this word has an entirely different meaning, namely 'a double bag, saddlebags' (similar to the Latin word *bisaccium*): South Italy, and Cappadocia. These two dialect enclaves lie at the extreme edges of the largest expansion of the Greek language, to the West and East respectively, and have been utterly out of touch with each other for more than a thousand years: so, the presence of this term in both areas is an indication of its early date.

The decision to include dialectal forms in the *ILNE*, has, as argued above, distinctive advantages, when it comes to tracing the history and the etymology of words, as well as the overall continuity of the language. In the words of Kahane and Kahane, when discussing the inclusion of dialectal material in the *Historical Dictionary of Modern Greek* (1992, 26): "The historical chain, progressing from phase to phase, which concentrated on the continuity of the culture, gave way to an exhibition of the Greek genius for language: the wealth of forms and meanings, of derivations and metaphorizations which a given lexeme underwent in the many dialects, turned into an impressive display of linguistic variation and creativity". However, it also entails a series of difficulties and additional problems, not commonly faced by compilers of standard historical dictionaries.

A first consequence is the greater mingling of synchrony and diachrony. Dialectal sources cover a considerable timespan, but are nevertheless viewed as belonging to the same synchrony in the *ILNE*[15]. This problem in the case of Greek is exacerbated by the existence of special historical circumstances in the early twentieth century, e.g. the Balkan wars (1912-13) resulting in considerable border expansion, and the Asia Minor disaster (1922) resulting in a complete population exchange between Greece and Turkey. These marked radical demographic and geographical changes, which affected the linguistic-dialectological picture of the area (Trudgill 2003, 47-48; Horrocks 2010, 428-437).

[15] In fact, this is standard practice in dialectal lexicography, cf. e.g. Penhallurick (2009) for English, Barbato & Varvaro (2004) for Italian and Katsouda (2012) for Greek.

The dialectal sources of the *ILNE* cover roughly 150 years, from the middle of the nineteenth century until the early twenty-first century, since the enrichment of the dialectal archive of the Dictionary is still ongoing, both on the basis of written sources and on the basis of annual *in situ* fieldwork with dialect speakers. As a result of the time-difference in the collection of data, the word forms present variation which might not only be dia-topical but also dia-chronic: different forms of the same word in the same variety, collected in different periods, and exhibiting an on-going change or conversely, exhibiting the retreat and obsolescence of the dialect.

5. Conclusion

This article has offered an overview of the main aims, problems and methodology of the *Historical Dictionary of Modern Greek* (*ILNE*), illustrated through specific examples. These theoretical and methodological issues have a bearing both on historical and on dialectal lexicography, as the *ILNE* is a rare type of dictionary combining both fields. We hope on the one hand to have shown how a solid methodology can lead to important and reliable results in the investigation of linguistic history, and on the other to have highlighted the main areas where the *ILNE* can serve as an example for facing and solving problems. We have not touched upon the most serious problem dogging all major national lexicographic projects, namely that of long duration and lack of adequate funding and personnel, but we believe that each successfully published volume is one step towards ensuring the viability of such projects world-wide.

Bibliography

Alexiou, Margaret. 1982. “Diglossia in Greece”. In *Standard Languages, Spoken and Written,* edited by William Haas, 156-192. Manchester: Manchester University.

Barbato, Marcello, and Varvaro, Alberto. 2004. “Dialect dictionaries”. *International Journal of Lexicography,* 17: 429-439.

Browning, Robert. 1982. “Greek diglossia yesterday and today”. *International Journal of the Sociology of Language,* 35: 49-68.

Buchi, Eva. 2016. “Etymological dictionaries”. In. *The Oxford Handbook of Lexicography,* edited by Philip Durkin, 338-349. Oxford: Oxford University Press

Chatzidakis (1905-1907) = Χατζιδάκις, Γεώργιος. 1905-1907. *Μεσαιωνικὰ καὶ Νέα Ἑλληνικά.* τ. Α΄ (1905) - τ. Β΄ (1907). Athens: P. D. Sakellarios.

Chatzidakis (1977) = Χατζιδάκις, Γεώργιος. 1977 [1926]. 'Διατὶ εἶμαι μὲν δημοτικιστὴς ἀλλὰ δὲν γράφω τὴν δημοτικήν' [Why I am a demoticist, but I do not write in the demotic]. In *Γλωσσολογικαί ἔρευναι* [Linguistic Researches] vol. 2, 358-370. Athens: Academy of Athens.

Considine, John. 1998. "Why do large historical dictionaries give so much pleasure to their owners and users?". In *EURALEX ' 98: Papers Submitted to the Eighth EURALEX International Congress on Lexicography*, edited by Thierry Fontenelle, Philippe Hilligsmann, Archibald Michiels, André Moulin and Siegfried Theissen, 579-587. Liège: University of Liège.

Cortelazzo, Manlio. 2007. *Dizionario Veneziano della lingua e della cultura popola nel XVI secolo.* Padova: La Linea Editrice.

Cramer, John Anthony. 1836. *Anecdota Graeca e codd. manuscriptis bibliothecarum Oxoniensium*, vol. 3, Oxford: Oxford University Press.

Crystal, David. 2003. *A Dictionary of Linguistics and Phonetics*. (5th edition). Oxford: Blackwell.

Dawkins, Richard.1916. *Modern Greek in Asia Minor. A study of the dialects of Sílli, Cappadocia and Phárasa with grammar, texts, translations and glossary.* Cambridge: Cambridge University Press.

Dollinger, Stefan. 2016. "National Dictionaries and Cultural Identity: Insights from Austrian, German, and Canadian English". In *The Oxford Handbook of Lexicography*, edited by Philip Durkin, 590-603. Oxford: Oxford University Press.

Drysdale Patrick D. 1989. "Etymological information in the general monolingual dictionary". In *Wörterbücher. Ein internationales Handbuch zu Lexikographie,* edited by Franz Josef Hausmann, Oskar Reichmann et al., vol. 1, 525-530. Berlin: De Gruyter.

Durkin, Philip. 2015. "Mackems, Geordies and ram-raiders: documenting regional variation in historical dictionaries". *English Language and Linguistics,* 19: 313-326.

Ferguson, Charles A., 1959. "Diglossia". *Word,* 15 (2): 325-340.

Fernandez, Mauro. 1995. "Los orígenes del término diglosia. Historia de una historia mal contada". *Historiographia Lingüística,* 22: 163-195.

Fortson, Benjamin W. 2004. *Indo-European Language and Culture*. Oxford: Blackwell.

Gaisford, Thomas. 1823. *Poetae minores Graeci*, vol. 2, Leipzig: Kühn.

Georgacas, Demetrios J., and Georgacas, Barbara 1991. "The Lexicography of Byzantine and Modern Greek". In *Wörterbucher. Ein internationales Handbuch zu Lexikographie,* edited by Franz Josef Hausmann, Oskar Reichmann, et al., vol. 2, 1705-1713. Berlin: De Gryuter.

Hawke, Andrew. 2016. “Quotation evidence and definitions”. In *The Oxford Handbook of Lexicography* edited by Philip Durkin, 176-206. Oxford: OUP.

Horrocks, Geoffrey C. 2010. *Greek: A History of the Language and Its Speakers*. Second ed. Oxford: Wiley-Blackwell.

Hudson, Alan. 2002. “Outline of a theory of diglossia”. *International Journal of the Sociology of Language,* 157: 1-48.

Karanastasis (1984-1992) = Καραναστάσης, Ἀναστάσιος. 1984-1992. *Ἱστορικὸν Λεξικὸν τῶν ἑλληνικῶν ἰδιωμάτων τῆς Κάτω Ἰταλίας* [Historical Dictionary of the Greek dialects of South Italy]. Athens: Academy of Athens.

Karasimos, Athanasios, Manolessou, Io and Melissaropoulou, Dimitra. 2020. “Creating a DTD template for Greek dialectal lexicography: the case of the Historical Dictionary of the Cappadocian dialect”. In *Proceedings of XIX EURALEX Congress: Lexicography for Inclusion*, Vol. I edited by Zoe Gavriilidou et al., 305-314, Democritus University of Thrace.

Kahane, Henry, and Kahane, Renée. 1992. “The Dictionary as Ideology. Sixteen case studies”. In *History, Languages and Lexicographers*, edited by Ladislav Zgusta, 19-76. Berlin: De Gruyter.

Katsouda (2012) = Κατσούδα, Γεωργία. “Διαλεκτική Λεξικογραφία: Επισκόπηση και Ζητήματα” [Dialect lexicography. Overview and Issues]. *Λεξικογραφικόν Δελτίον,* 26: 77-159.

Katsouda (2016) = Κατσούδα, Γεωργία. “Ο νέος τόμος του Ιστορικού Λεξικού της Νέας Ελληνικής της Ακαδημίας Αθηνών: συγχρονικές προοπτικές” [The new volume of the Historical Dictionary of Modern Greek: synchronic perspectives]. *Studies in Greek Linguistics*, 36: 151-160.

Konstantinidou and Tzamali (2012) = Κωνσταντινίδου Μαγδαληνή and Αικατερίνη Τζαμάλη. 2012. “Οι πηγές του Ιστορικού Λεξικού της Ακαδημίας Αθηνών”. [The sources of the Historical Dictionary of the Academy of Athens] *Λεξικογραφικόν Δελτίον*, 26: 279-382.

Kostakis (1986-1987) = Κωστάκης, Θανάσης. 1986-1987. *Λεξικό της Τσακωνικής διαλέκτου*. [Dictionary of the Tsakonian dialect]. Athens: Academy of Athens.

Koukoules, Phaidon. 1933. “The Athens Modern-Greek Lexicon”. *Journal of Hellenic Studies*, 53 (1): 1-8.

Koumanoudis (1900) = Κουμανούδης, Στέφανος.1900. *Συναγωγὴ νέων λέξεων ὑπὸ τῶν λογίων πλασθεισῶν ἀπὸ τῆς Ἁλώσεως μέχρι τῶν καθ’ ἡμᾶς χρόνων* [Compilation of new words, coined by learned scholars since the Fall of Constantinople until our times]. 2 vols. Athens: P.D. Sakellarios.

Mackridge, Peter. 1990."Katharevousa (c. 1800-1974): An Obituary for an Official Language". In *Background to Contemporary Greece,* edited by Marion Sarafis and Martin Eve, vol. 1, 25-51. London: Merlin Press/ Savage, Maryland: Barnes & Noble.

Mackridge, Peter. 2009. *Language and National Identity in Greece, 1766-1976.* Oxford: Oxford University Press.

Manolessou (2012) = Μανωλέσσου, Ιώ. 2012. "Το *Ιστορικόν Λεξικόν της Νέας Ελληνικής* ως Ιστορικό Λεξικό". [The "*Historical Dictionary of Modern Greek*" as a historical dictionary] *Λεξικογραφικόν Δελτίον* 26: 11- 75.

Manolessou, Io. 2014. "Learned Byzantine language from the viewpoint of linguistics". In *The Language of Byzantine Learned Literature*, edited by Martin Hinterberger, 13-33. Turnhout: Brepols.

Manolessou, Beis and Bassea-Bezantakou (2012) = Μανωλέσσου, Ιώ, Σταμάτης Μπέης and Χριστίνα Μπασέα-Μπεζαντάκου. 2012. "Η φωνητική απόδοση των νεοελληνικών διαλέκτων και ιδιωμάτων". [The phonetic representation of the Modern Greek dialects]. *Λεξικογραφικόν Δελτίον,* 26: 161- 221

Merkin, Reuven. 1983. "Historical Dictionaries". In *Lexicography: Principles and Practice,* edited by Reinhard Rudolf Karl Hartmann, 123-133. London: Academic Press.

Papadopoulos (1958-1961) = Παπαδόπουλος, Ἄνθιμος. 1958-1961. *Ἱστορικὸν Λεξικὸν τῆς Ποντικῆς Διαλέκτου*. [*Historical Dictionary of the Pontic dialect*]. Athens: Epitropi Pontiakon Meleton.

Penhallurick, Rob. 2009. "Dialect dictionaries". In *The Oxford History of English Lexicography,* edited by Anthony Philip Cowie, vol. II, 290-313. Oxford: Clarendon Press.

Pernot, Hubert. 1897. *Grammaire grecque moderne*. Paris: Garnier.

Reichmann, Oskar. 1990a. "Formen und Probleme der Datenerhebung I: Synchronische und diachronische historische Wörterbücher". In *Wörterbucher. Ein internationales Handbuch zu Lexikographie,* edited by Franz Josef Hausmann, Oskar Reichmann et al., vol. 2, 1588-1611. Berlin: De Gryuter.

Reichmann, Oskar. 1990b. "Das gesamtsystembezogene Wörterbuch". In *Wörterbucher. Ein internationales Handbuch zu Lexikographie,* edited by Franz Josef Hausmann, Oskar Reichmann et al., vol 2, 1391-1416. Berlin: De Gruyter.

Reichmann, Oskar. 2007. "Stand und Aufgaben der historischen Lexikographie des Deutschen". *Lexicographica,* 23: 199-230.

Reichmann, Oskar. 2012. *Historische Lexikographie. Ideen, Verwirklichungen, Reflexionen an Beispielen des Deutschen, Niederländischen und Englischen.* Studia Linguistica Germanica 111. Berlin/Boston: De Gruyter.

Schiefer, Lieselotte, and Pompino-Marschall, Bernd.1996. "Phonetische Transkription". In *Schrift und Schriftlichkeit. Writing and its Use. Ein interdisziplinäres Handbuch internationaler Forschung,* edited by Hartmut Günther, Otto Ludwig, et al., 2. Halbband, 1583-1591. Berlin/New York: de Gruyter.

Trudgill, Peter. 2003. "Modern Greek dialects. A preliminary classification". *Journal of Greek Linguistics,* 4: 45-84.

Tseronis, Assimakis. 2002. "Diglossic past and present lexicographic practices. The case of two Greek dictionaries". *Language Problems & Language Planning,* 26: 219-252.

Wiegand, Herbert Ernst. 1998[2]. "Historische Lexikographie". In *Sprachgeschichte, Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung,* edited by Werner Besch et al., 643-715. Berlin: De Gruyter.

Vagiakakos 1977 = Βαγιακάκος, Δικαῖος: *Γεώργιος Ν. Χατζιδάκις (1848-1941). Βίος καὶ ἔργον* [Georgios N. Chatzidakis (1848-1941). Life and works]. Athens: Academy of Athens.

Dictionaries and corpora

DARE = Cassidy, Frederic, and Houston Hall, Joan. 1985-2012. *Dictionary of American Regional English,* vol. 5. Cambridge, MA: Belknap Press of Harvard University Press.

DWB = Grimm, Jacob, and Grimm, Wilhelm. 1838-1961. *Deutsches Wörterbuch*, 17 vol., Leipzig. Stuttgart: S. Hirzel.

EDD = Wright, Joseph. 1898-1905. *English Dialect Dictionary,* vol. 6. Oxford: Henry Frowde.

HNC = Hellenic National Corpus. (Institute of Language and Speech Processing)
http://corpus.ilsp.gr/index.php?current_page=main&lang=en

ILNE= Ἱστορικὸν Λεξικὸν τῆς Νέας Ἑλληνικῆς, τῆς τε κοινῶς ὁμιλουμένης καὶ τῶν ἰδιωμάτων, τ. *1-6 (Α- διάλεκτος) [Historical Dictionary of Modern Greek, both of the Standard and the dialects, vol. 1-6, A-dialect].* Athens: Academy of Athens.

ILNE (2021) = *Ἱστορικὸν Λεξικὸν τῆς Νέας Ἑλληνικῆς, τῆς τε κοινῶς ὁμιλουμένης καὶ τῶν ἰδιωμάτων,* 7ος τ., *τεῦχ. Α, διάλεξη-δόγης*

[Historical Dictionary of Modern Greek, both of the Standard and the dialects, vol. 7, part 1, dialeksi-dogis]. Athens: Academy of Athens.

ILNE-MR = Κανονισμός Συντάξεως του Ιστορικού Λεξικού της Νέας Ελληνικής [Manual of Regulations of the Historical Dictionary of Modern Greek]. Athens: Academy of Athens, 2011.

TLG = Thesaurus Linguae Graecae, California, University of Irvine. http://stephanus.tlg.uci.edu/

OED = Murray, James. 1884–1928 *Oxford English Dictionary*. Oxford University Press.

SAOB = Svenska Akademiens ordbok: Ordbok över svenska språket utgiven av Svenska Akademien. Stockholm, 1898-.

WNT = Woordenboek der Nederlandsche Taal (1864-1998) / Electronic edition 1999. Rotterdam: AND Publishers b.v.

CHAPTER ELEVEN

THE IMPORTANCE OF THE BIBLE TRANSLATION FOR THE OLD CZECH LEXICOGRAPHY[1]

KATEŘINA VOLEKOVÁ

Abstract

This article examines the Old Czech Bible translation in the Late Middle Ages and its importance for historical lexicography. During the fourteenth and fifteenth centuries, separate translations of Psalter and Gospel lectionary and four distinct versions of translation of the complete Bible, the so-called redactions, emerged, surviving in more than one hundred manuscripts and incunabula. The Bible has a very significant place in the historical lexicology and lexicography of the Old Czech language for several reasons. Translated from the Latin source text, the Bible is the most extensive and lexically rich text and one of the earliest sources in the vernacular. Due to numerous translations from different periods, it can be used as a language development indicator as well.

In this article, desirable improvements in the research of the Old Czech biblical vocabulary are discussed; in particular, more critical editions of the particular Old Czech Bible translations are needed to complete the existing excerption from several selected biblical sources that were prepared for the *Old Czech Dictionary* in the 1960s. Nowadays, this need is filled by a project focused on publishing the digital editions of individual biblical books from representative texts of Old Czech Bible translation.

---

[1] The preparation of this study was financed within the statutory activity of the Czech Language Institute of the Czech Academy of Sciences (RVO No. 68378092). It has also used data provided by the LINDAT/CLARIAH-CZ Research Infrastructure (https://lindat.cz), supported by the Ministry of Education, Youth and Sports of the Czech Republic (Project No. LM2018101).

1. Introduction

The Old Czech translation of the Bible provides priceless linguistic evidence for the study and description of the vernacular language. The first attempts to translate the Latin Vulgate appeared during the reign of the last Přemyslids at the turn of the thirteenth and fourteenth centuries, when other Czech literary works such as songs, legends and chronicles were also composed for the first time. For the following centuries, the translation of the biblical text remained one of the most common literary activities and therefore greatly influenced development of the literary language. In this article, after briefly introducing the history of Bible translation into Old Czech and various source texts, I will discuss why the Old Czech Bible is so important for developing the historical lexicography and lexicology of the Czech language. Then I will present critical editions of the fourteenth- and fifteenth-century Old Czech biblical texts that have been published so far. Presently, digital editions of the Old Czech Bibles are being prepared, which scholars can use for the Old Czech vocabulary analysis and which lexicographers in particular can use when writing an Old Czech dictionary.

1.1 The history of translation of the Bible into Old Czech

The Old Czech Bible translation started with glosses, vernacular notes in the margin or between the lines of a text to translate a difficult word or an obscure passage from the Holy Writ. The Jagić's glosses from the Viennese manuscript are the oldest vernacular notes in the Latin biblical text. The partly Old Slavonic, partly Old Czech glosses were written in the Latin Gospels of Matthew and Mark at the beginning of the twelfth century by a monk expelled from the Sázava monastery where the Slavonic liturgy was cultivated until it was banned in 1097.[2] This precious language material contains Czech words from everyday life (e.g. *kvas* “ferment”, *múka* “flour”, *plena* “diaper, drape”) as well as abstract words (e.g. *nelacno* “difficult, hardly”, *prázný* “empty, idle”, *osúditi* “decide, judge”).

The vernacular translation of continuous biblical text emerged approximately two centuries later, at a time when literature written in the Czech language began to develop in the Kingdom of Bohemia. In the first instance, the selected sections of the Bible used in the liturgical practice were translated into Old Czech, namely a Book of Psalms at the end of the thirteenth century and Gospel lections at the beginning of the fourteenth

[2] The glosses are written on the fols. 241r-249r of the so-called Radobibel, Vienna, Österreichische Nationalbibliothek, shelfmark Cod 1190; cf. Vintr 1986a, 77-113, Sichálek 2017, 67.

century. The Old Czech Psalter was originally intended for private study of the Bible in monasteries for the needs of the nuns, later it was used by noblewomen for their personal piety as well. The translator of the Psalms did his job responsibly: he managed to find appropriate Czech words for Latin expressions without creating unnecessary neologisms.[3] The first Gospel lectionary in Czech was made for preachers to facilitate the preparation of homilies in the vernacular.

The first Czech translation of the whole Bible was finished in the 1350s during the reign of King Charles IV, when Bohemia became the center of Central European culture and education. It was an important literary, cultural and social achievement and a successful attempt to convert a demanding Latin text. The great flourishing of Czech literature during the second half of the fourteenth century necessitated a revision of the biblical translation which manifested itself in the adaptation of the manner of expression to contemporary Czech, e.g. replacing obsolete idioms with modern words (e.g. the substitution of verbs *hýřiti* "to sin", *pohýřiti* "to make a mistake" for verbs *přěstúpiti, přěstupovati, protiviti sě* etc.) and the stabilization of biblical terminology by promoting one of several synonymous terms (e.g. the term *spravedlnost* for Latin *iustitia* "justice" instead of *pravedlnost, právo* or *pravda*). The Hussite reform movement in the fifteenth century stimulated an extraordinary interest in the Holy Writ in vernacular; the biblical text has therefore been revised and newly translated into Old Czech several times. By the end of the fifteenth century, there were four distinct versions of the translation, the so-called redactions. The second and third redactions of the Old Czech Bible translation come from the first decades of the fifteenth century, from the period of pre-Hussite reform efforts, and the fourth redaction was prepared for the first printed Czech Bible in the 1480s.

Moreover, various retellings of the biblical stories can be assigned to biblical texts as well, such as a Czech translation of the popular *Historia scholastica* written by French theologian Petrus Comestor (d. 1178) which surprisingly emerged in the Czech lands after the first Old Czech Bible translation at the end of the fourteenth century, whereas in most European literatures the vernacular Bible is rather preceded by a translation of Petrus

[3] Cf. Kyas 1962, 9; Kyas 1997, 32. The translator was probably a man lecturing at a Prague Chapter school. However, it is very likely due to the audience (nuns and noblewomen) that the biblical manuscripts were later copied and modified by women as well; cf. Vintr 1998, 219. The first Czech Psalter translation is preserved in its entirety only in the late copy in a manuscript of the Bibliothek des Evangelischen Predigerseminars in Wittenberg from the mid-fourteenth century (shelfmark A VI 6).

Comestor's *Historia scholastica* or similar works retelling the biblical narrative.[4]

The Old Czech Bible translation has been preserved in a large number of written sources, more than one hundred manuscripts including 27 complete Bibles, 26 Old Testaments, 41 New Testaments, 15 Psalters, and 10 Lectionaries. Nearly another hundred fragments of varying length survived from the fourteenth and fifteenth centuries, and as well as that, the biblical text was printed seven times by the end of the fifteenth century (see Fig.11-1).[5] No two copies are exactly the same: scribes often improved text while copying and, at their own discretion, they replaced words with updated or more appropriate expressions.

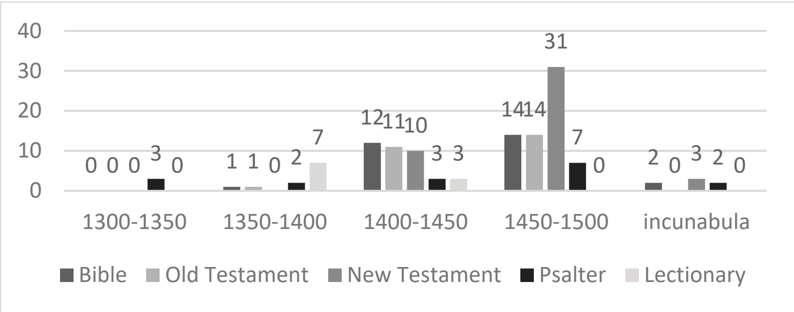


Fig. 11-1: Numbers of the Old Czech Bible manuscripts and prints.

The vernacular translation of the complete Holy Writ or its parts provides priceless linguistic evidence for the study and description of the language; as a result, the Old Czech Bible has a very significant place in the historical lexicology and lexicography of Czech. In my opinion, there are four main reasons why this is so. The first reason is the date of the first Bible translations into Old Czech. The earliest literary works in Old Czech were preserved in written sources in the late thirteenth century. The oldest surviving Czech texts were mostly religious songs, Czech translations of Latin legends of saints, epics, chronicles, and adaptations of chivalrous romances, all in verse. Therefore, the earliest translations of Psalter and Gospel lectionary from the beginning of the fourteenth century belong among the first prosaic texts in vernacular language. And due to the overall low number of vernacular source texts, the Old Czech Bible manuscripts

[4] Cf. Kyas 1997, 64-65; Sichálek 2017, 81.
[5] For the recent list of the biblical manuscripts and prints, see Voleková and Svobodová 2019, 518-532.

from the first half of the fourteenth century form a substantial part of all sources used in the illustrative quotations in the *Staročeský slovník* [*Old Czech Dictionary*] (1968-2008), the historical dictionary describing Czech vocabulary from the first written source texts until 1500, which comprises the entries beginning with letters N to P (see Fig. 11-2).[6]

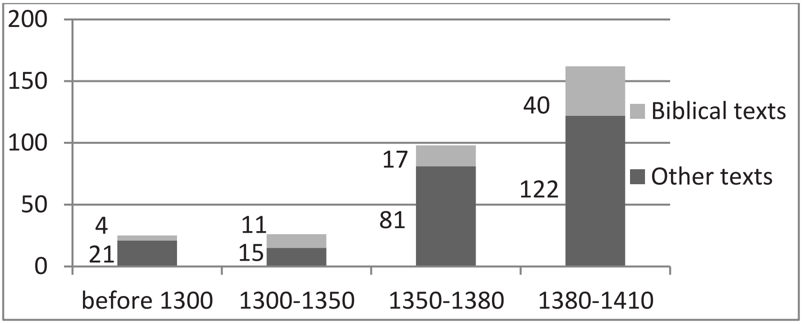


Fig 11-2: The manuscript sources of the *Old Czech Dictionary* (1968-2008).

The quotations from the biblical source texts are essential for the *Elektronický slovník staré češtiny* [*Electronic Dictionary of Old Czech*] (2006-) as well.[7] It was originally conceived as a dictionary without examples, but this kind of historical dictionary was found to be inadequate, because it did not provide proof of a word's actual existence and use in the medieval period. Therefore, from 2020, examples are added to the entries beginning with *Přib-*, specifically two illustrative quotations for every meaning, if possible, the oldest one and a representative one.[8] The biblical source texts often meet these criteria; the quotations are often chosen from the earliest Old Czech Psalters and Lectionaries, or from one of the four biblical redactions, so the importance of the Old Czech Bible translation for historical lexicography is even greater than before.

The frequent choice of a biblical quotation is also due to the fact that the first complete Old Czech Bible from the 1350s is the most extensive and

[6] The basic sigla of the manuscript sources are counted in the graph according to the *Staročeský slovník. Úvodní stati, soupis pramenů a zkratek*, 1968. The historical dictionaries of Old Czech are introduced in more detail in Chapter 12 by Irena Fuková and Markéta Pytlíková.

[7] "Elektronický slovník staré češtiny," Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka, accessed 23 June 2021, https://vokabular.ujc.cas.cz/hledani.aspx.

[8] See Fuková and Pytlíková, Chapter 12 in this volume.

lexically rich text from this period and this also applies to the following redactions of the Old Czech Bible translation. The Bible is a collection of various writings, most variable in terms of genres (narration, lyric poem, prophecy, letter, wisdom book, etc.), and it describes various topics; some of them are quite special, for example religious rites or prophetic visions. The complete Bible is also one of the longest texts; the first redaction of the Old Czech Bible contains almost half a million tokens and seven thousand lemmas.[9] Therefore, it contains priceless linguistic evidence of the usage of numerous lexemes from the centre and the periphery of the Old Czech vocabulary. The biblical texts provide us with the first usage of some lexemes (e.g. *načřieti* "refill, fill up", *zpěvákyni* "female singer"), archaic and obsolete words (e.g. *náv* "world of the dead, underworld", *rozščutina* "tearing of clothes, hole", *stkmieti* "to accumulate assets"), new words (e.g. *bratřivie* "brotherhood", *nájmišče* "a place where workers were hired"), or new meanings of lexemes (e.g. *obřězánie* "circumcision", *přělúditi* "to knock down"). Many Czech words, especially archaisms, remain restricted only to the biblical texts and do not appear in any other texts (e.g. *náručník* "string instrument held in hand", *posdravitedlný* "cleansing, redeemable", *snízanie* "humming, grumbling", *vzochviti sě* "to focus on", *zlatohlavník* "person dressed in a cloth with gold threads").

The third reason is that the text in the Old Czech Bible has been translated from the Latin Vulgate,[10] therefore a Latin equivalent usually helps to uncover the meaning of the Old Czech expression. Of course, there is certain variability in the dependence on the Latin source text—although the Bible is considered as an authoritative and sacred text, some translations were looser, primarily trying to produce a periphrastic meaning, even at the cost of deviating slightly from the source text, while other translations were rather literal, translating the text word by word. The first translation of the whole Bible belongs predominantly to a group of meaning-to-meaning translations trying to put the Scriptures into more understandable wording.[11] Also some modifications of the Czech biblical translation belong to this group, especially the one that appeared in the Utraquist environment in the

[9] Based on the automatic lemmatization of the corpus manager and text analysis software "Sketch Engine", Lexical Computing CZ s.r.o., accessed 10 January 2020, https://www.sketchengine.eu.

[10] The direct Latin source texts of the individual Old Czech redactions are unknown, they were unspecified manuscripts with medieval versions of Vulgate, especially with the so-called Paris Bible.

[11] Cf. Kyas 1997, 44; Sichálek 2017, 76-77.

second half of the fifteenth century.[12] The Utraquist translators, presumably the priests and preachers, targeted the average reader; they were more concerned with clarity and intelligibility of the biblical text than its literalness. Other translations were close and literal, e.g. the first Psalter, which originated in the Czech glosses written between the lines of the Latin Psalms,[13] or the third redaction of the whole Bible, which strictly followed the Latin text even at the expense of clarity of translation.[14] At all times, translators have tried to find adequate means of language expression. But a possible influence of the source language on the target language should be always kept in mind. The original Latin word formation was sometimes reflected in Old Czech words without considering the principles of Czech word formation and the meaning of Czech prefixes and suffixes. A significant part of the neologisms appearing only in the Old Czech Bible translation are loan translations from the Latin (e.g. *nadzbytečný* “considerably larger than sufficient” from Latin *superaubundans* or *supereffluens*, *přinalezovánie* “decision, intentional act” from Latin *adinventio*).[15]

Finally, the great advantage for modern lexicographers of historical Czech is that in four various translations and other smaller revisions of the Old Czech Bible, we can follow and compare equivalents and their phonology, morphology, and valency during the two hundred years of the Old Czech period, i.e. from the beginning of Czech literature to the peak of medieval literary production at the end of the Middle Ages. Over the two centuries, biblical texts have both reflected linguistic changes and stimulated them. Regarding lexicology, lexical formation processes and changes in Old Czech vocabulary can be traced in the biblical texts very well. For example, we follow how one lexeme was replaced by another. The lexeme *napajadlo* “a place with the water to feed cattle” from Song 4:2 was used only in the oldest Bibles, represented by the Dresden Bible from the second half of the 1460s[16] and the Olomouc Bible from 1417.[17] It is a loose translation capturing the interpretive meaning of the biblical verse. However,

[12] The Utraquists (from the Latin sub utraque specie, meaning "in both kinds") were a moderate faction of the Hussites. After the defeat of the more radical Taborites and Orphans, they eventually reunited with the pope and the Catholic Church, and they were allowed to offer Communion under both kinds (both as bread and wine) officially.

[13] Cf. Kyas 1997, 32; Sichálek 2017, 69.

[14] Cf. Kyas 1997, 100.

[15] The total number cannot be calculated due to heterogeneous processing of the Old Czech vocabulary in the modern dictionaries.

[16] The manuscript was destroyed in 1914, formerly stored in Dresden, Saxon State and University Library, shelfmark O e 85.

[17] The manuscript is stored in Olomouc, Research Library, shelfmark M III 1/I-II.

the younger texts from the first half of the fifteenth century, the Litoměřice Bible,[18] representative text of the second redaction, and the Padeřov Bible,[19] representative text of the third redaction, chose the equivalent *umyvadlo* "a washbasin, a place to wash" that is closer to the Latin word *lavacrum*, and the printed Prague Bible,[20] representative text of the fourth redaction, changed it for the word *kupadlo* "a place to bathe" with a similar meaning, but more appropriate in the context.

> *zubi tvoji jako stáda střižených ovec, ješto jsú vzešly z napajadl* Dresden Bible, Song 4:2 (similarly Olomouc Bible, *z umyvadla* Litoměřice Bible, Padeřov Bible, *z kupadla* Prague Bible)
> *dentes tui sicut greges tonsarum quae ascenderunt de lavacro* Vulgate[21]
> *thy teeth as flocks of sheep, that are shorn which come up from the washing* Douay-Rheims Bible[22]

There were many similar cases of searching for the appropriate equivalent. For example, the Old Czech equivalent for Latin adjective *dyscolus* "incorrigible" from 1 Pet 2:18 differs almost in each translation: *nekáný* "incorrigible", *poteklý* "stray, deviant", *poběhlý* "stray, deviant", *zlý* "bad, evil".

> *služebníci, poddáni buďte ve všiej bázni pánóm netolik dobrým a rovným, ale také i nekáným* Dresden Bible, 1 Pet 2:18 (*poteklým* Olomouc Bible, Padeřov Bible, *poběhlým* Prague Bible, *zlým* Olomouc Lectionary)
> *servi, subditi estote in omni timore dominis, non tantum bonis et modestis, sed etiam dyscolis* Vulgate
> *servants, be subject to your masters with all fear, not only to the good and gentle, but also to the froward* Douay-Rheims Bible

Various translation strategies of specific biblical lexemes (flora, fauna, clothes, etc.) have been applied by translators. For example, in the case of

---

[18] The manuscript is now missing, formerly Litoměřice, State Regional Archives, Episcopal Collections Litoměřice, shelfmark BIF 2.
[19] The manuscript is stored in Vienna, Austrian National Library, shelfmark Cod. 1175.
[20] [*Bible pražská*]. Prague: [Printer of the Prague Bible], 1488. GW 4323, ISTC ib00620000. The used exemplar is stored in Vienna, Austrian National Library, shelfmark Ink 13.C.5.
[21] "The Clementine Vulgate Project," accessed 10 January 2020, http://vulsearch.sourceforge.net.
[22] "Douay–Rheims Bible Online," accessed 29 September 2020, http://www.drbo.org/index.htm.

exotic and mythical animals from Ps 90:13 (Latin *aspis*, *basilisus*, *draco*), pan-Slavic names with a more general meaning are used in the Bible translations (Old Czech *had* “snake”, *jěščer* “reptile”, *san* “snake, dragon”, *zmek* “creeping animal”), or new words with specific meaning are created (Old Czech *králík* “poisonous snake with a crown”, *ostrovid* “animal with sharp eyes”, *zřieček/řieček* “animal killing with its gaze”), or Latin terms have been adapted into Czech (*aspid*, *bazilišek*, *drak, draksan*).[23]

> *na tak řečeném hadu aspidu i po baziliškovi choditi budeš i potlačíš leva i zemka* Wittenberg Psalter, Psalms 90:13 (*králíkovi ... zřiečku ... ještěra* Museum Glossed Psalter, *králíkovi ... řiečku ... ještěra* Dresden Bible, *hadem aspidem ... basiliškem ... draksana* Clementine Psalter, *bazilišku ... aspidu ... san* Chapter Psalter, *králíkovi ... bazilišku ... ještěra* Poděbrady Psalter, *aspidu ... bazilišku ... draka* Olomouc Bible, *králíkovi ... ostrovidu ... draka* Litoměřice Bible, *hadě ... bazilišku ... draka* Boskovice Bible)
> *super aspidem et basiliscum ambulabis, et conculcabis leonem et draconem* Vulgate
> *thou shalt walk upon the asp and the basilisk: and thou shalt trample under foot the lion and the dragon* Douay-Rheims Bible

One of the processes in the Old Czech vocabulary is the inconsistency and gradual stabilization of meanings in word-forming suffixes that caused high polylexy in Old Czech. We can often record cases of a fluctuation in the word formation, for example Old Czech adjectives *nebytedlný*, *nebydlitedlný*, *nepřěbyvatedlný* for Latin *inhabitabiblis* “uninhabitable”.[24] But in some cases, the Czech equivalent is gradually stabilizing, for example the Czech lexemes for a servant. The old expression *panošě* was replaced at the beginning of the fifteenth century by a contemporary common word *sluha*, which remained in the biblical translation until the first printed Czech Bible, which preferred the synonym *služebník*.

> *nenie učenník nad svým mistrem ani panošě nad svým pánem* Dresden Bible, Matt 10:24 (*sluha* Olomouc Bible, Litoměřice Bible, Padeřov Bible, *služebník* Prague Bible)
> *non est discipulus super magistrum, nec servus super dominum suum* Vulgate
> *the disciple is not above his master, nor the servant above his lord* Douay-Rheims Bible

---

[23] Cf. Chládková and Nejedlý 2007, 148-152.
[24] Cf. the phrase *terra inhabitabilis* from Jer 2:6, 17:6, 48:9 in *Staročeský slovník*, s.v. nebydlitedlný.

2. Desirable improvements for the research of the Old Czech biblical vocabulary

The full use of the Old Czech biblical texts for vocabulary processing is limited due to their insufficient familiarity to modern scholars. Individual versions of the Old Czech Bible translations, their language and style have not yet been fully described either. During the processing of the Old Czech historical dictionary, lexicographers from the Department of Language Development of the Czech Language Institute of the Czech Academy of Sciences work with the two existing index card catalogs of citations from Old Czech texts, which also include lexical examples from the biblical sources. The first excerption was done for the first Old Czech grammar and dictionary written by Jan Gebauer at the turn of the nineteenth and twentieth centuries.[25] The contemporary diplomatic editions of the oldest sources (e.g. Psalters from the fourteenth century) and some selected biblical manuscripts (e.g. Litoměřice-Třeboň Bible, Kladruby Bible, Pernštejn Bible) were included in the excerption. But it was selective, not covering all major revisions and redactions. Later, this lexical excerption was found to be insufficient, when new critical editions were prepared and the knowledge of individual redactions of the Old Czech Bible translation was improved. Therefore, a new excerption including several selected biblical sources was prepared for the *Old Czech Dictionary* in the 1960s, but it also did not cover all important texts.[26] As a consequence, present-day lexicographers still have to list variants from selected representatives of Old Czech Bible translation themselves to complete the existing excerption. It is not easy work; it is time-consuming, they need to work with manuscripts or their photocopies, because most of the sources are not available in modern editions.

Therefore, lexicographers are looking for a way in which the processing of the dictionary can be simplified when using important biblical examples. A complementary excerption from biblical manuscripts is no longer possible, it was rejected because of its disadvantage — the time-consuming task does not produce adequate results. And so instead of excerption, an editorial processing of biblical texts is more useful and nowadays digital editions are a better solution, because they are multifunctional and they can

[25] "Lístková kartotéka Jana Gebauera," Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka, accessed 23 June 2021, https://vokabular.ujc.cas.cz/kartoteky.aspx?db=2.

[26] *Lístková kartotéka Staročeského slovníku* [*Index Card Catalogue of the Old Czech Dictionary*], stored in the Department of Language Development of the Czech Language Institute of the Czech Academy of Sciences.

be used for different types of linguistic research, using corpus tools. However, there is only a limited number of modern transcribed editions of the Old Czech biblical texts that can be converted into a digitized form. Only critical editions of the earliest Old Czech translations from the fourteenth century have been prepared so far, the first Gospel lectionary and Psalter translations by Josef Vintr (Vintr 1977, Vintr 1986b) and the first Bible translation by Vladimír Kyas (Kyas 1981, Kyas 1985, Kyas 1988, Kyas and Kyasová and Pečírková 1996, Pečírková et al. 2009).[27]

More critical editions of the Old Czech Bible translations, especially from the fifteenth century, are needed, so the editorial processing of new digital editions of Old Czech Bibles has been included among tasks of the textological-editorial section of the Department of Language Development. Based on the practical experience of how long it takes to prepare an edition of one whole Bible due to its extraordinary length, editors decided to process the text by single books. Therefore, they did not work on the edition of another entire Bible of the second or the third redaction, but started to compile and edit one biblical book from various Bible translations selected from all surviving Old Czech manuscript and print biblical sources.[28] The digital editions of the Old Czech Bibles are used in several databases operated by the Czech Language Institute. The web portal *Vokabulář webový* [*Web Vocabulary*] provides access to a portal module *Edice* [*Editions*] with the full texts of primary sources in continuous form[29] and to a corpus *Staročeská textová banka* [*Old Czech Text Bank*], supplemented with gradual lemmatization and morphological tagging.[30] The other

[27] The last volume was prepared by a team led by Jaroslava Pečírková.

[28] The preliminary work began with the most frequent text, with the Gospel of Matthew, and in 2019, the digital editions from the eight chosen representative texts of the Old Czech translations from the mid-fourteenth century to the end of the fifteenth century were published. In 2020, the publication of the Old Czech Gospel of Mark from the same sources was completed. Editors are currently working on the digital editions of Luke Gospel and the Book of Psalms, the most common and widespread book of the Old Testament. The edition of the Old Czech Psalms is supported in the years 2020-2022 by the project of the Czech Science Foundation No. 20-06229S *The First Printed Old Czech Psalter (a philological analysis and critical edition)*. Then editors will continue with the Gospel of John, the estimated date of publication is 2024.

[29] "Edice," Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka, accessed 23 June 2021, https://vokabular.ujc.cas.cz/moduly/edicni/.

[30] "Staročeská textová banka," Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka, accessed 23 June 2021,
https://vokabular.ujc.cas.cz/banka.aspx?idz=STB, or

representation of the Old Czech biblical texts is provided in a newly launched database *Diabible (Czech Bible Translation in Diachrony: A Database of Late Medieval Biblical Texts)*.[31] Inspired by an electronic concordance of the fifteenth-century translations of the New Testament into Old Polish,[32] the database *Diabible* was created on similar principles, enabling study of the Old Czech Bible translation by means of a comparison of individual biblical verses.

3. Conclusion

The historical lexicographers of the Czech Language Institute benefit significantly from close collaboration with the editors of the Old Czech and Early Modern Czech literary texts from the same department. The editions, including gradually also various Old Czech Bible translations, are accessible to scholars, students and the general public in the databases *Web Vocabulary*, *Old Czech Text Bank*, and *Diabible*; each database provides a different approach to the study of language. One of the aims of editing Old Czech Bible translations is to use them more intensively for semantic exemplification in the *Electronic Dictionary of Old Czech* as these texts often offer the oldest and best examples of the Old Czech lexemes and their meanings.

Publishing the digital editions of all biblical books from chosen representative manuscripts and prints of the Old Czech Bible translation will surely take a while; so far, one biblical book per year has been processed. Nevertheless, the new *corpus* of Old Czech Bibles can be used by lexicographers and for linguistic research on language development before it is completed. More detailed analysis will certainly bring more knowledge about the changes in biblical vocabulary and formation of the biblical style, which will contribute to the research of Old Czech lexicology in general.

---

https://korpus.vokabular.ujc.cas.cz/first_form?corpname=STB-1.1.18.1&usesubcorp=.

[31] "Diabible," Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka, accessed 23 June 2021, http://diabible.com.

[32] "Piętnastowieczne przekłady Nowego Testamentu," Instytut Języka Polskiego PAN, accessed 23 June 2021, http://stnt.ijp.pan.pl/.

Bibliography

Chládková, Věra, and Nejedlý, Petr, 2007. "Ze staročeského slovníku (Mezi exotickými a bájnými tvory staročeského žalmisty)." In *Listy filologické*, 130: 148-152.

Kyas, Vladimír. 1962. *Česká předloha staropolského žaltáře*. Praha: ČSAV.

Kyas, Vladimír, ed. 1981. *Staročeská Bible drážďanská a olomoucká: kritické vydání nejstaršího českého překladu Bible ze 14. století. I, Evangelia*. Praha: Academia.

Kyas, Vladimír, ed. 1985. *Staročeská Bible drážďanská a olomoucká s částmi Bible litoměřicko-třeboňské: kritické vydání nejstaršího českého překladu Bible ze 14. století. II, Epištoly, Skutky apoštolů, Apokalypsa*. Praha: Academia.

Kyas, Vladimír, ed. 1988. *Staročeská Bible drážďanská a olomoucká: kritické vydání nejstaršího českého překladu Bible ze 14. století. III, Genesis - Esdráš*. Praha: Academia.

Kyas, Vladimír, Kyasová, Věra, and Pečírková, Jaroslava, eds. 1996. *Staročeská Bible drážďanská a olomoucká: kritické vydání nejstaršího českého překladu bible ze 14. století. IV, Tobiáš - Sirachovec*. Paderborn: Schöningh.

Kyas, Vladimír. 1997. *Česká bible v dějinách národního písemnictví*. Praha: Vyšehrad, Řím: Křesťanská akademie.

Pečírková, Jaroslava, et al., eds. 2009. *Staročeská Bible drážďanská a olomoucká s částmi Proroků rožmberských a Bible litoměřicko-třeboňské: kritické vydání nejstaršího českého překladu Bible ze 14. století. V/1, Izaiáš - Daniel, V/2, Ozeáš - 2. Makabejská*. Praha: Academia.

Sichálek, Jakub. 2017. "European Background: Czech Translations." In *The Wycliffite Bible: Origin, History and Interpretation*, edited by Elizabeth Solopova, 66-84. Leiden, Boston: Bril.

Staročeský slovník. Úvodní stati, soupis pramenů a zkratek. 1968. Praha: Academia.

Staročeský slovník 1968–2008. Praha: Academia.

Vintr, Josef, ed. 1977. *Die älteste tschechische Evangeliare*. München: Verlag Otto Sagner.

Vintr, Josef. 1986a. "Die tschechisch-kirchenslavischen Glossen des 12. Jahrhunderts in der Bibel Sign. 1190 der Nationalbibliothek in Wien (sog. Jagić-Glossen)." In *Wiener Slavistisches Jahrbuch*, 32: 77-113.

Vintr, Josef, ed. 1986b. *Die älteste tschechische Psalterübersetzung*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Vintr, Josef. 1998. "České biblické slovo jako vývojový faktor stylové stratifikace národního jazyka – od Jagićových glos 12. století po moderní ekumenický překlad." In *Pocta 650. výročí založení Univerzity Karlovy v Praze. 1. díl: Sborník příspěvků přednesených zahraničními bohemisty na mezinárodním sympoziu v Praze 20.–26. srpna 1998*, edited by Alexandr Stich and Otakar Mališ, 215-228. Praha: Univerzita Karlova.
Voleková, Kateřina, and Svobodová, Andrea, eds. 2019. *Staročeské biblické předmluvy*. Praha: Scriptorium.

Online databases and sources

"Diabible," Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka, accessed 23 June 2021, http://diabible.com.
"Douay–Rheims Bible Online," accessed 29 September 2020, http://www.drbo.org/index.htm.
"Edice," Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka, accessed 23 June 2021, https://vokabular.ujc.cas.cz/moduly/edicni/.
"Elektronický slovník staré češtiny," Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka, accessed 23 June 2021, https://vokabular.ujc.cas.cz/hledani.aspx.
"Lístková kartotéka Jana Gebauera," Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka, accessed 23 June 2021, https://vokabular.ujc.cas.cz/kartoteky.aspx?db=2.
"Piętnastowieczne przekłady Nowego Testamentu," Instytut Języka Polskiego PAN, accessed 23 June 2021, http://stnt.ijp.pan.pl/.
"Sketch Engine," Lexical Computing CZ s.r.o., accessed 10 January 2020, https://www.sketchengine.eu.
"Staročeská textová banka," Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka, accessed 23 June 2021, https://vokabular.ujc.cas.cz/banka.aspx?idz=STB, or https://korpus.vokabular.ujc.cas.cz/first_form?corpname=STB-1.1.18.1&usesubcorp=.
"The Clementine Vulgate Project," accessed 10 January 2020, http://vulsearch.sourceforge.net.

CHAPTER TWELVE

HISTORICAL ONLINE DICTIONARY AS A DYNAMIC PROCESS: FOURTEEN YEARS AND THREE STAGES OF THE *ELECTRONIC DICTIONARY OF OLD CZECH*[1]

IRENA FUKOVÁ AND MARKÉTA PYTLÍKOVÁ

Abstract

The *Elektronický slovník staré češtiny [Electronic Dictionary of Old Czech]* (2006–) is an online historical dictionary describing the documented vocabulary of the Old Czech period (from the twelfth century to 1500). It is a team effort in progress, initiated in 2005 by scholars of the Czech Language Institute of the Czech Academy of Sciences and following up on the extensive *Staročeský slovník* [*Old Czech Dictionary*] (1968-2008).

This article introduces the *Electronic Dictionary of Old Czech*, including its historical background, and describes three stages of its development in the fourteen years of its existence. Within the description of the third stage, we focus on the current change concerning the content and form of the dictionary itself and the individual entries, especially on the decision that the described lexemes will be newly documented by short examples from Old Czech texts. We introduce the basic points of the chosen solution and present the main principles that have been established for the selection of examples.

The development of the modern online dictionary of Old Czech is presented as a dynamic process of finding a balance between the requirement

[1] This study was prepared in the Czech Language Institute of the Czech Academy of Sciences and supported by LINDAT/CLARIAH-CZ, a project of the Czech Ministry of Education, Youth and Sport No. LM2018101. The study uses data of the website Vokabulář webový (https://vokabular.ujc.cas.cz), provided by the research infrastructure LINDAT/CLARIAH-CZ (https://lindat.cz).

for quick results and the requirement for the most accurate processing and presentation of the vast preserved Old Czech language material.

1. Introduction

The *Elektronický slovník staré češtiny* [*Electronic Dictionary of Old* Czech] is an online historical dictionary, a team effort in progress, initiated in 2005 by scholars from the Department of Language Development of the Czech Language Institute of the Czech Academy of Sciences and published since 2006 on the web portal *Vokabulář webový [Web Vocabulary]* (https://vokabular.ujc.cas.cz). The aim of the dictionary is to describe the documented vocabulary of the Old Czech period (from the twelfth century to 1500).

This article introduces the *Electronic Dictionary of Old Czech* within its historical background, describes three stages of its development in the fourteen years of its existence and focuses on the current change in working principles, which lie primarily in the decision to use examples from Old Czech texts in entries from *přib*- to *Ž*.

2. Historical background of the *Electronic Dictionary of Old Czech*

The origin and development of the *Electronic Dictionary of Old Czech* is closely related to the history of the description of Old Czech vocabulary in the last one hundred years.[2] The first scientific work consistently treating the vocabulary of Old Czech was the *Slovník staročeský* [*Dictionary of Old Czech*] by Jan Gebauer, published in 1903-1916. The dictionary was based on using index cards prepared by the author, his students and collaborators to record excerpted illustrative quotations from Old Czech texts. The description of the meaning and polysemic structure of the lexical units was brief, but the dictionary contained a large number of examples from Old Czech texts as well as from texts of the sixteenth century. The headword was spelled in the form assumed around the year 1300; this tradition was later adapted by more modern dictionaries of Old Czech. Unfortunately, the author's death left his great work unfinished; the dictionary includes only lexical units from letter *A* to the beginning of letter *N* (the last lexical unit is the noun *netbalivost*).

Until the middle of the twentieth century, attempts to follow up on Gebauer's dictionary were unsuccessful. During the 1950s, preparations

[2] For detailed information regarding the history of description of Old Czech vocabulary, see Homolková 2007.

began for a modern dictionary of Old Czech within the Czech Language Institute (*Ústav pro jazyk český*) of the Czechoslovak Academy of Sciences (*Československá akademie věd*); in particular an extensive excerption from hundreds of Old Czech manuscripts and incunabula, resulting in a vast index card catalog of citations with more than one million cards. In 1968, a first volume of *Staročeský slovník* [the *Old Czech Dictionary*] was published, summarizing its underlying principles, methodology and sources.[3] The first issue of the dictionary followed almost immediately, starting with the lexical unit *na*. The *Old Czech Dictionary* thus ideologically as well as practically followed the work of Jan Gebauer.[4]

The *Old Czech Dictionary* was based on a thorough theoretical background and detailed working principles in accordance with structuralist linguistic theory.[5] The main principles were summarized, especially in the introductory articles: the headword was spelled in the form assumed around the year 1300, the meaning, valence and collocability of lexical units were systematically described along with the extensive presentation of examples from Old Czech texts. Entries were always supervised by a semantics editor and a graphemics editor.

Twenty-six issues of the dictionary were published between 1968 and 2008, covering lexical units from *na* to *při.* Unfortunately, the insufficient staffing of the lexicographic team for such a demanding dictionary work in the last decade of the twentieth century, resulting in a significant slowdown in the work on the *Old Czech Dictionary*, and the need to describe the Old Czech vocabulary in its entirety in a reasonable time frame led in 2005 to the decision to terminate the *Old Czech Dictionary* in its current form (cf. Homolková 2007, 233-234).

At the same time, the content and form of the macro- and microstructure of a dictionary that would better meet the time and staff requirements (along with new publishing options) were investigated. The discussion took into account the state of the description of the Old Czech vocabulary at the time of termination of the *Old Czech Dictionary*; the vocabulary of Old Czech was then lexicographically described and documented with examples in the

[3] *Staročeský slovník. Úvodní stati, soupis pramenů a zkratek* 1968.

[4] Lexical units of the first part of the letter N (*na* - *netbalivost*) are therefore present in both dictionaries and offer interesting material for comparing the two lexicographical approaches.

[5] Igor Němec, the chief editor of semantics of the *Old Czech Dictionary*, was the author of the essential studies on Old Czech vocabulary, based on the lexicological research: *Vývojové postupy české slovní zásoby [Developmental Procedures of Czech Vocabulary]* (Němec 1968) and *Rekonstrukce lexikálního vývoje [Reconstruction of Lexical Development]* (Němec 1980).

alphabetical range *A–při*. The Old Czech lexical units from *přib-* to *Ž* were then practically unexplored (cf. Table 12-1). Basic information about this part of the alphabet was provided only by *Malý staročeský slovník [Little Old Czech Dictionary]* (Bělič, Kamiš and Kučera 1978), a very simple historical dictionary, intended primarily for university teaching purposes. This dictionary covered the whole alphabet, but very selectively; only the words no longer existing in Modern Czech or having significantly different meaning were described (Bělič, Kamiš and Kučera 1978, 5), with only a basic explanation of the meaning and without examples.

	A–M	*N–při*	*přib–Ž*
Slovník staročeský [Dictionary of Old Czech] by Jan Gebauer (1903-1916)	brief description, with examples	–	–
Staročeský slovník [Old Czech Dictionary] (1968-2008)	–	complex description, with examples	–
Malý staročeský slovník [Little Old Czech Dictionary] (1978)	only selected lexical units, brief description, no examples	only selected lexical units, brief description, no examples	only selected lexical units, brief description, no examples

Table 12-1: Description of the Old Czech vocabulary at the time of terminating the Old Czech Dictionary (2005).

3. *The Electronic Dictionary of Old Czech*

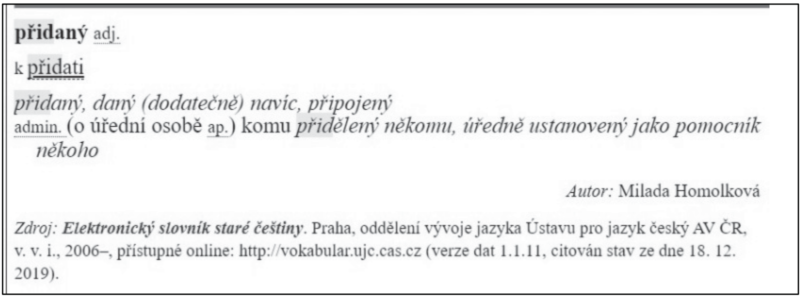
The result of the discussion within the Institute of Czech Language was a new project, the *Electronic Dictionary of Old Czech*.[6] Its concept was based on the *Old Czech Dictionary*, but significantly simplified; the expected benefits were acceleration of the processing as well as more flexible publishing via the web (cf. Vajdlová 2011, 29; Černá 2011, 40).[7]

[6] The head of the project was M. Vajdlová (2005-2018) and Š. Šimek (since 2019).
[7] At the same time, it was decided to process electronically and to publish all three other Old Czech dictionaries via *Vokabulář webový* (vokabular.ujc.cas.cz), i.e. the *Slovník staročeský* by Jan Gebauer (published via the web in 2007), the *Little Old*

The *Electronic Dictionary* has undergone three stages in its fourteen years of existence. This dynamic process is an attempt at finding a balance between the requirement for quick results and the requirement for the most accurate processing of the vast preserved Old Czech language material.

3.1 Stage 1 of the Electronic Dictionary of Old Czech: 2005-2010

In Stage 1 (from 2005 to 2010), the *Electronic Dictionary* aimed to complete the processing of the Old Czech vocabulary from *přib-* to *Ž* as quickly as possible so that the final part of the alphabet would be covered over the next five years. This short processing time was an absolute priority. Therefore, in this part of alphabet, the *Electronic Dictionary* processed only those lexical units that were not yet lexicographically described in the *Little Old Czech Dictionary*. Furthermore, the description of the lexical units on all language levels was greatly simplified; for example, word formation data were given only selectively. The entries did not go through editorial unification and did not contain examples from Old Czech texts (cf. entry *přidaný* in Fig. 12-1). During Stage 1, 17,500 entries were processed (Vajdlová 2011, 29).



přidaný adj.

k přidati

přidaný, daný (dodatečně) navíc, připojený
admin. (o úřední osobě ap.) komu *přidělený někomu, úředně ustanovený jako pomocník někoho*

Autor: Milada Homolková

Zdroj: ***Elektronický slovník staré češtiny***. Praha, oddělení vývoje jazyka Ústavu pro jazyk český AV ČR, v. v. i., 2006–, přístupné online: http://vokabular.ujc.cas.cz (verze dat 1.1.11, citován stav ze dne 18. 12. 2019).

Fig. 12-1: The adjective *přidaný* 'added; attached; assigned' processed during Stage 1 of the *Electronic Dictionary of Old Czech* and published via *Vokabulář webový*. There is no numbered hierarchy of meanings and no examples from Old Czech texts. (Source: https://vokabular.ujc.cas.cz/hledani.aspx 18. 12. 2020)

The aim was to give the basic description of those Old Czech lexical units which were not included in the *Little Old Czech Dictionary*. Together,

Czech Dictionary (published via the web in 2006) and the *Old Czech Dictionary* (published via the web in 2008-2011). The web site *Vokabulář webový* allows searching all these dictionaries together (including the *Electronic Dictionary of Old Czech*) and comparing the results.

these two lexicographic works provided a basic overview of the documented Old Czech words within this part of alphabet. The comparability was facilitated by the simultaneous electronic processing of the *Little Old Czech Dictionary*; both sources were available on the website *Vokabulář webový* and could be searched together (cf. Fig. 12-2).

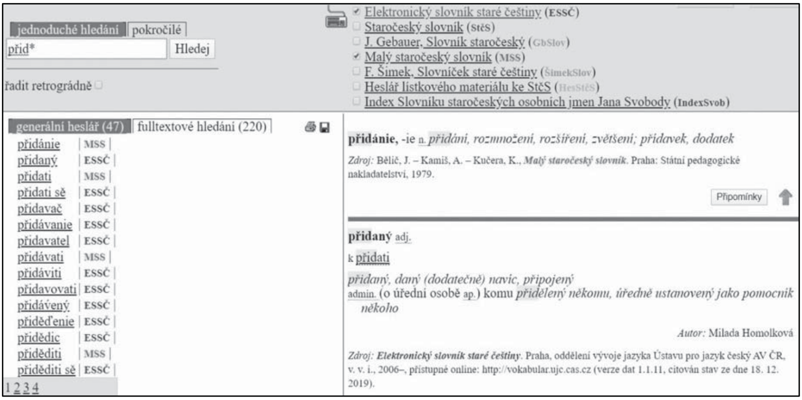


Fig. 12-2: On the left, there are search results of the lexical units beginning with *přid-* from both dictionaries together, sorted alphabetically (ESSČ—*Electronic Dictionary of Old Czech*; MSS—*Little Old Czech Dictionary*). On the right there are first two lexical units of the search from two different sources: the noun *přidánie* 'addition; widening; increase; supplement' from the *Little Old Czech Dictionary*, and the adjective *přidaný* 'added; attached; assigned' from Stage 1 of the *Electronic Dictionary of Old Czech*, already described in Fig. 12-1 (Source: https://vokabular.ujc.cas.cz/hledani.aspx 18. 12. 2019)

However, during Stage 1, the difference between the conceptual principles of the *Little Old Czech Dictionary* and the *Electronic Dictionary of Old Czech* became more apparent, bringing significant difficulties and sometimes even distortions of the overall picture of the Old Czech vocabulary. The difference was most distinct in word families, where some lexical units of a word family were already present in the *Little Old Czech Dictionary* (usually the base words), and some were processed in the *Electronic Dictionary of Old Czech* (often the derivatives).[8]

[8] There were also differences in headword spelling. In the *Little Old Czech Dictionary*, the form of the headword assumed around 1300 was not reconstructed; instead, the form (or forms) documented in source texts was used (cf. Bělič, Kamiš and Kučera 1978, 6). The *Little Old Czech Dictionary*, unlike the *Electronic*

For example, the adjective *učitedlný* was published in the *Little Old Czech Dictionary* with the meaning ‘effective, powerful’. The derived noun *učitedlnost*, not present in the *Little Old Czech Dictionary*, was then processed in the *Electronic Dictionary* with the correct meaning ‘docility, understanding’, which did not correspond to the meaning of the adjective *učitedlný*, from which the noun *učitedlnost* was derived. Inspection of the material sources revealed that the meaning of the base word in the *Little Old Czech Dictionary* was erroneous and should be ‘capable of learning, insightful; capable of teaching’ and infrequently also ‘eager, diligent’.

The above-mentioned problems have led to some changes in the form and content of the *Electronic Dictionary* that have been applied in the subsequent part of the alphabet (from letter *A*); in the first place, the interconnection of the *Electronic Dictionary* and the *Little Old Czech Dictionary* was abandoned.

3.2 Stage 2 of the Electronic Dictionary of Old Czech: 2010-2015

In Stage 2 (from 2010 to 2015), lexical units in the alphabetical range *A* to *M* (over 14,000 entries) were processed in the *Electronic Dictionary of Old Czech*. All documented Old Czech words from *A* to *M*[9] were described in the *Electronic Dictionary of Old Czech*, no matter whether they were also documented in other historical dictionaries (i.e. in the *Dictionary of Old Czech* by Jan Gebauer and the *Little Old Czech Dictionary*).

Language description of lexical units in Stage 2 was more detailed than in Stage 1, including word formation data and more elaborated description of meaning. For the entries of the letter *M*, the unifying semantic editing was tested.

By 2015, the *Electronic Dictionary of Old Czech* and its predecessor the *Old Czech Dictionary* had together finally described almost the entire documented Old Czech vocabulary, with the exception of the part from *přib-* to *Ž*, where it was also necessary to use the *Little Old Czech Dictionary* (cf. Table 12-2).

Dictionary of Old Czech, also sometimes processed the words with the same word root and the same meaning but with different suffixes together in one entry.

[9]I.e. documented in the index card catalogue of citations..

	A–M	*N–při*	*přib–Ž*
Slovník staročeský [Dictionary of Old Czech] by Jan Gebauer (1903-1916)	brief description, with examples	–	–
Staročeský slovník [Old Czech Dictionary] (1968-2008)	–	complex description, with examples	–
Electronický slovník staré češtiny [Electronic Dictionary of Old Czech] (2006-)	**ESSČ stage 2** all lexical units, detailed description, no examples	–	**ESSČ stage 1** only the lexical units **not** included in the *Little Old Czech Dictionary*, brief description, no examples
Malý staročeský slovník [Little Old Czech Dictionary] (1978)	only selected lexical units, brief description, no examples	only selected lexical units, brief description, no examples	only selected lexical units, brief description, no examples

Table 12-2: Description of the Old Czech vocabulary after Stage 2 of the *Electronic Dictionary of Old Czech* (2015).

3.3 Stage 3 of the Electronic Dictionary of Old Czech: 2016-

The problematic interconnection of two very different historical dictionaries, the *Electronic Dictionary of Old Czech* and the *Little Old Czech Dictionary*, in the part from *přib-* to *Ž* as well as the too brief and simplified manner of processing entries in this part of the alphabet in the *Electronic Dictionary of Old Czech* led in 2016 to Stage 3. Its aim has been to rework and complete entries from *přib-* to *Ž* in a consistent, systematic and more detailed manner, completely independent of the *Little Old Czech Dictionary*.

For this stage, some major changes in the working principles have been implemented:[10]

- all the documented lexical units are processed, including those already present in the *Little Old Czech Dictionary*;

[10] The principles of the concept have been summarized in internal guidelines (Šimek at al. 2019).

- grammatical data and word-forming data are described systematically, description of meaning is detailed;
- in this part of the alphabet, the entries are systematically unified by semantics and graphemics editors;
- every meaning of a lexical unit is documented by two examples (if possible) from Old Czech texts, including the oldest one.

The first part of the lexical units, 450 entries (from *Přib* to *přidržovati sě* and from *přímĕnek* to *připásati*), was published via the web site Vokabulář webový in April 2020 and the change in concept is clearly visible by comparing Figs. 12-2 and 12-3. The noun *přidánie*, formerly present only in the *Little Old Czech Dictionary*, and the adjective *přidaný*, published during Stage 1 in a very simple manner, has been in Stage 3 newly processed according to the new working principles. There is a complex description of lexical meaning, including numbered hierarchy; each meaning is supported by illustrative examples taken from Old Czech texts.

přidánie, -ie n.

k přidati

1. čeho *přidání, dodání, dodatečné připojení navíc k něčemu:* addicio přidánie KlarGlosA 2290 (De iudice); s přidáním přidšenie, točížto skrze chi řecká [jména] mají čtena býti VýklHebrL 232r aspiratione addita
   jur. **přidánie let** komu k čemu *prohlášení někoho nezletilého za plnoletého:* přidání mi let k tomu vkladu DeskyMorOl 28 (1481)
2. *přídavek, dodatek k něčemu:* věrnému našemu za vděčné služby … takovú věc dáváme neb takového přidánie pójčujeme tiemto listem PrávHorM 50r praerogativam; aby s ní [tvrzí] učinili [majitelé] …, což by se jim zdálo najpodobnějie …, s tiemto dalším přidáním: že těch dvě stě kop gr. … jest rozděleno mezi strany svrchupsané ArchČ 26,319 (1428)
3. čeho (majetku) *převedení, postoupení:* vzal sem list toho kúpenie …, otpovědi a vóle přidánie BiblOl Jr 32,11 (předánie BiblDrážď) rata; tu ves … našim potomkóm poostavujem …, naše přidánie té vsi … nemuože býti oblášče vyplacováno ArchČ 16,128 (1454)

K 2: sr. StčS předánie[1]

Zdroj: ***Elektronický slovník staré češtiny***. Praha, oddělení vývoje jazyka Ústavu pro jazyk český AV ČR, v. v. i., 2006–, přístupné online: http://vokabular.ujc.cas.cz (verze dat 1.1.15, citován stav ze dne 29. 10. 2020).

---

přidaný adj.

k přidati

1. *přidaný, daný (dodatečně) navíc, připojený:* Levi přidaný nebo přispořený nebo přiložený anebo přijatý nebo přistavený VýklHebrL 210v additus; jiné přímluvy sú nové, přidané ChelčPost 84v
2. admin. (o úřední osobě ap.) komu *přidělený někomu, úředně ustanovený jako pomocník někoho:* proti takovému [protiprávně jednajícímu člověku] máme sobě všichni pomoci vedle rozkázanie zprávce a rady jemu přidané ArchČ 3,260 (1426); tiem vládni paní Anna s poručníky sobě přidanými vedle kšaftu Andrlíkova ListářPlz 1,412 (1446)

Zdroj: ***Elektronický slovník staré češtiny***. Praha, oddělení vývoje jazyka Ústavu pro jazyk český AV ČR, v. v. i., 2006–, přístupné online: http://vokabular.ujc.cas.cz (verze dat 1.1.15, citován stav ze dne 29. 10. 2020).

Fig. 12-3: The noun *přidánie* and the adjective *přidaný* processed during Stage 3 and published via Vokabulář webový. (Source: https://vokabular.ujc.cas.cz/hledani.aspx 29. 10. 2020.)

4. Examples in Stage 3 of the *Electronic Dictionary of Old Czech* (from *přib-* to *Ž*)

The main goals of adding examples for Old Czech lexical units from *přib-* to *Ž* in the *Electronic Dictionary* have been providing evidence of the existence of a lexical unit and its specific meaning as well as proving its age. Another significant reason was that Old Czech lexical units in the part of the alphabet *přib-* to *Ž* are those least documented from source texts in other historical dictionaries (see Tables 12-1 and 12-2).[11]

The question was in what number and in which form should the examples be selected. The detailed working principles of the *Old Czech Dictionary*, requiring sometimes four or more examples per meaning[12], were not applicable; a minimalist solution was sought to achieve the goals and still maintain a relatively short processing time. The part of the internal guidelines governing the selection and presentation of the Old Czech examples in the *Electronic Dictionary of Old Czech* (Šimek et al. 2019) were therefore based on the *Old Czech Dictionary* concept, but in a reduced and simplified manner–not in the means of analyzing excerpted citations (the analysis of lexical unit is still based on all documented citations), but rather in the means of the process of their selection for presentation.

4.1 Two Old Czech examples for every meaning of a lexical unit

The guidelines state that each meaning of a lexical unit should ideally be documented with two examples: The first, a compulsory one, is the oldest one; the second example should be "a representative one, lexicographically valuable, well-documented, and exemplifying the described meaning of the word". This example should also cover as many of the following criteria as possible: exemplification of central valency; collocability; the dominant context; and genre.

In the following examples, we use the verb *přibližovati sě*[13] to demonstrate how variable valency or collocability of one meaning is covered by two examples.[14]

[11] The examples for the lexical units from *A* to *M* are documented in the *Slovník staročeský* by Jan Gebauer; the part from *N* to *při* is richly documented in the *Old Czech Dictionary*.
[12] *Staročeský slovník. Úvodní stati, soupis pramenů a zkratek* 1968, 41-44.
[13] https://vokabular.ujc.cas.cz/hledani.aspx, s.v. *přibližovati sě*, 29. 10. 2020.
[14] In the following examples we highlight the discussed part of a citation in bold and translate the citation to English in the commentary. For the translation of the Bible citations, we use King James Bible Standard text of 1769

> 1. (k) komu/čemu, proti komu ap. *přibližovat se, dostávat se do blízkosti něčeho:* v ohlavi a v úzdě líci jich [koně a mezka] stiehni, již **nepřibližijí sě k tobě** ŽaltWittb 31,9 approximant ad te; **nepřibližij sě [synu] ke dveřóm** jejieho [nevěstčina] domu BiblOl Pr 5,8 ne appropinques foribus

In Sense 1 of the verb *přibližovati sě* 'to approach, to get close to something' is the dominant valency and collocability "to whom, to what". The valency/collocability "to whom" is documented by the oldest example from the *Wittenberg Psalter* (ŽaltWittb): '*[horse or mule] whose mouth must be held in with bit and bridle lest they **come near unto thee***'; the other collocability "to what" is documented by the second example from the *Olomouc Bible* (BiblOl): '***come not nigh the door** of her [strange woman's] house'*.

In the following example, different collocability is demonstrated. Sense 7 of the verb *přibližovati sě* is 'to occur, to come'.

> 7. (o časovém úseku, události ap.) *přibližovat se, (téměř) nastávat, nadcházet:* aj, toť **sě dni přibližijí**, praví hospodin BiblDrážď Jr 7,32 venient; již **sě smrt přibližuje**, aj, rucě mé oslaběly sú OrlojK 438

A subject of this verb is either a time period when something happens or the event itself. This different collocability is covered by two examples as well: the oldest example from the *Dresden Bible* (BiblDrážď) represents the verb with the time period, i.e. "days", as a subject:

> '*Therefore, behold, **the days come**, saith the Lord*.'

The second example from the Old Czech translation of the Suso's *Horologium Sapientiae* (OrlojK) was chosen to document the use of the verb with an event, in this case "death", as a subject:

> *'**death is coming**, behold, my hands are weakened'.*

(https://www.kingjamesbibleonline.org/); the citations from non-biblical manuscripts are translated by authors.

4.2 One Old Czech example for a specific use of a lexical unit within a meaning, especially for an idiomatic, figurative or terminological use

For a specific use of a lexical unit within a sense (shift in a meaning, figurative use, terminological use, etc.), one more example is given; a phrase or an idiom is documented by one example as well.[15] For instance, within the sense 1 of the verb (see below) *přičiniti*[16] 'put something close to something', documented by quotations from the Old Czech translation of Petrus Comestor's *Historia Scholastica* (ComestC) and *The Franciscan Medicine Book* (LékFrantA), we also give an example for a slightly different meaning, 'to lay hand on someone with the intention to hurt him' (BiblOl): '*then came they, and* ***laid hands*** *on Jesus, and took him'*, and also one example for the idiom "put one's hand", meaning 'to help' from *The Exposition of the Apocalypse* by Jacob of Mies (JakZjev): '*when God* ***put his hand*** *at a certain hour..., then all things are going well*'.

1. co k čemu/komu *přiblížit, přiložit, dát do těsné blízkosti něčeho:* k tomu ohni když ruku přičiníš, horkost čiješ ComestC 12r adhibes; přičiň ucho k té dieře a vylezeť [škvor] hned LékFrantA 46v; co (ruku) k komu *vztáhnout na někoho:* tehdy přistúpichu [vojáci po Jidášově polibku] a **rucě přičinichu** k Ježíšovi i jěchu jej BiblOl Mt 26,50 manus injecerunt in Iesum
přičiniti ruku svú *napomoci něčemu:* když buoh **přičiniti ráčí ruku svou** v jistou hodinu …, tehdy všecky věci dobře se vedou JakZjev 321r

Within sense 1 of the verb *přibósti*[17] 'to pin something to something; to attach something to something by a sharp object' (ComestC), we also present an example for a figurative use—**"pierce my flesh with your fear"** (BiblOl).

1. co/koho (k) čemu čím *přibodnout, připevnit k něčemu zabodnutím něčeho:* provazy a háky meděnými přibodniechu [kůže] zemi obvázány ComestC 74r affixis; vrh [Saul] kopím, … mnieše Davida, … by jej k stěně přibodl ComestC 146v configere – fig.: **přiboď**

[15] Every specific use of a lexical unit within a sense has an explicit definition in modern Czech, except for a figurative use, indicated only by an abbreviation *fig.* (see *přibósti* 1).
[16] https://vokabular.ujc.cas.cz/hledani.aspx s.v. *přičiniti* 29. 10. 2020
[17] https://vokabular.ujc.cas.cz/hledani.aspx s.v. *přibósti* 29. 10. 2020

strachem tvým tělo mé BiblOl Ps 118,120 (sepni ŽaltWittb, zboď ŽaltPod, skroť ŽaltKap, sevři BiblLit, probodni BiblPad) confige

5. Conclusion

An essential aspect of the new conceptual guidelines (including the guidelines for the presentation of examples) is the completion of Stage 3 of *Elektronický slovník staré češtiny* [the *Electronic Dictionary of Old Czech*] in the near future. That is why the guidelines strive for maximum brevity and simplicity and create conditions for less time-consuming processing of entries, compared to the *Old Czech Dictionary*. At the same time, however, they try to fulfil the ideas of the meaning and mission of a historical dictionary intended primarily for scholars and professionals, so that the *Electronic Dictionary of Old Czech* can be an adequate completion of unfinished parts of the *Old Czech Dictionary*. In the future, these two modern lexicographical works should accurately describe the Old Czech vocabulary as a whole, as seen in Table 12-3.

	A–M	*N–při*	*přib–Ž*
Staročeský slovník [Old Czech Dictionary] (1968-2008)	–	complex description, with examples	–
Electronický slovník staré češtiny [Electronic Dictionary of Old Czech] (2006-)	**ESSČ stage 2** detailed description, no examples	–	**ESSČ stage 3** detailed description, with examples

Table 12-3: Description of the Old Czech vocabulary after the completion of Stage 3 of the *Electronic Dictionary of Old Czech*.

Bibliography

Bělič, Jaromír, Kamiš, Adolf, and Kučera, Karel. 1978. *Malý staročeský slovník*. Praha: Státní pedagogické nakladatelství.

Elektronický slovník staré češtiny. 2006-. Praha: Ústav pro jazyk český AV ČR. http://vokabular.ujc.cas.cz.

Černá, Alena. 2011.“Vokabulář webový—internetová platforma ESSČ (jeho současnost a budoucnost).” *Jazykovědné aktuality,* 48 (1-2): 40-43.

Šimek, Štěpán at al. 2019. *Návod ke zpracování ESSČ: Výběrové pracovní zásady pro zpracování ESSČ v abecedním rozmezí při- až Ž jako stručného dokladového redigovaného slovníku* (Unpublished internal guidelines of the Department of Language Development).

Gebauer, Jan. 1903-1916. *Slovník staročeský 1-2*. Praha: Česká grafická akc. společnost Unie.

Homolková, Milada. 2007. "Historická lexikografie." In *Kapitoly z dějin české jazykovědné bohemistiky*, edited by Jana Pleskalová et al., 199-242. Praha: Academia.

Němec, Igor. 1968. *Vývojové postupy české slovní zásoby*. Praha: Academia.

Němec, Igor. 1980. *Rekonstrukce lexikálního vývoje*. Praha: Academia.

Staročeský slovník. 1968-2008. Praha: Academia.

Staročeský slovník. Úvodní stati, soupis pramenů a zkratek. 1968. Praha: Academia.

Vajdlová, Miloslava. 2011. "Minulost a budoucnost Elektronického slovníku staré češtiny." *Jazykovědné aktuality,* 48 (1-2): 29-34.

CHAPTER THIRTEEN

LATIN AND ITALIAN INFLUENCE ON CROATIAN DICTIONARIES UP TO JAKOV MIKALJA'S *THESAURUS LINGUAE ILLYRICAE*[1]

ANA MIHALJEVIĆ

Abstract

In this article, the author analyzes foreign influence in bilingual and multilingual dictionaries and word lists containing a Latin and a Croatian column dating from the early modern period up to the middle of the seventeenth century. The analysis begins with the first word lists, focusing on the first Croatian dictionary, Vrančić's *Dictionarium quinque nobilissimarum Europae linguarum* (1595), and concludes with Mikalja's *Thesaurus Linguae Illyricae* (1649/1651). These multilingual dictionaries are an important testimony of the linguistic world and offer information useful for the study of the social and cultural history of the languages involved (Verbeke 2015, 29). In them, two languages are in explicit and continuous interaction (Demo 2011). These dictionaries are regarded as a place of language contact and contact-induced syntactic change. The influence of Latin and Italian (due to the dictionaries being mostly compiled by compilers educated in Italy, fluent in Italian, and probably mirroring certain Italian dictionaries) on the syntactic structure of Croatian is analyzed. In compiling a dictionary of a non-standard language, which Croatian was at

[1] This work has been fully supported by the Croatian Science Foundation under the project Retro-digitization and Interpretation of Croatian Grammar Books before Illyrism IP-2018-01-3585. Research group: dr. Petra Bago, dr. Martina Kramarić, dr. Ivana Lovrić Jović, dr. Ana Mihaljević, dr. Sanja Perić Gavrančić, dr. Ivo Pranjković, dr. Diana Stolac, dr. Ljiljana Šarić, dr. Barbara Štebih Golub, dr. Tamara Tvrtković. Project leader: dr. Marijana Horvat. The project is conducted by the Institute of Croatian Language and Linguistics in Zagreb.

the time, it was deemed necessary to form new words and constructions, with the Croatian dictionary column mirroring the Latin and/or Italian columns. Word-formation was sometimes the only way to fill the empty space in the Croatian column of a dictionary. The dictionary authors have sometimes filled in these blanks through loan translations or by adopting loan words.

1. Introduction

The beginnings of Croatian lexicography have been described in Musulin 1959, Franolić 1983, Filipović 1986, Samardžija 2011, Considine 2017, etc.[2] In this article, I will focus on Latin and Italian influence on the development of Croatian lexicography from the first Croatian dictionary to the *Thesaurus Linguae Illyricae* written by Jakov Mikalja[3]. In essays on the history of the Croatian language and lexicography, it is often stated that these two languages had a strong influence on Croatian, but the scope of that influence has not yet been researched in detail.

2. Latin vs. the vernacular

In Neo-Latin literature and literature on the development of national and vulgar languages, it is often stressed that there was an ongoing debate lasting several centuries on whether one should write in Latin or the vernacular.[4] The most famous text dating from the beginning of this debate, Dante Alighieri's *De vulgari eloquentia*, was "a manifesto propagating the use of the vernacular but published in Latin in order to be understood by a wide audience" (Verbeke 2015, 27). At approximately the same time, supporters of Latin defended the use of Latin as a literary and scholarly language whilst others propagated the use of the vernacular. Participants in this debate included Leon Battista Alberti (1404-1472)—*Della famiglia*, *Momus sive de principe;* Joachim Du Bellay (ca. 1522-1560)—*Deffence et illustration de la langue francaise* (1549); Erycius Putaneus (1574-1646)—*Iuventutis Belgiae laudatio* (1607), Jakob Grimm (1785-1863)—*De desiderio patriae* (1830), Romolo Quirino Amaseo (1489-1552), Carlo Sigonio (1522/3-1584), Johann Gottfried Herder (1744-1803) Arthur Shopenhauer

[2] For a history of Slavic lexicography see Derksen 2019.
[3] Mikalja 2011.
[4] For more details on this debate see Verbeke 2015. For the relation between Latin and the vernacular in the early modern period see Schaffenrath and Winkler 2018.

(1788-1860), etc. This discussion lasted well into the nineteenth century (De Santis 1995; IJsewijn and Sacre 1993 as cited in Verbeke 2015).[5]

The importance of Latin throughout Europe is well-known and generally accepted; its influence and importance in Croatia are also well-known. The range and depth of Latin's influence on Croatian, however, has still not been sufficiently researched. The relation between Latin and the Croatian dialects is even more complicated. Latin was the official language in Croatia until 1847, the latest of any nation, due to various political and social factors. That does not mean, however, that all Croats knew or understood Latin, but it is one of the factors behind the fruitful Neo-Latin production in Croatia.[6] "In Croatia, for over a thousand years, the Latin language was used for belles-lettres, education, and scholarship, for politics and religion, for private correspondence and public administration, at times addressing the Croatian public, at times an international audience." (Jovanović 2017, 35). Neo-Latin literature encompasses around 7000 titles. In Croatia, there was another literary language mostly in use from the ninth to the seventeenth centuries—Croatian Church Slavonic. This language was written in Glagolitic script and was mostly used in liturgy and religious texts.[7] It is a higher literary style within the Medieval Croatian diasystem, which encompassed Croatian and Croatian Church Slavonic diglossia.[8] It developed from Old Church Slavonic, created by Saints Cyril and Methodius for translating biblical, liturgical, religious, literary, and administrative

[5] Writing about the situation in Netherlands Anne Dykstra states: "The seventeenth- and eighteenth-century international audience for Dutch academic publications apparently still had an adequate command of Latin, says Roelevink (ibid [1990, 35]), considering the fact that authors and publishers continued to make frequent use of the language. In the eighteenth century the active knowledge of Latin, according to Roelevink (1990, 36), eroded much quicker than its passive use. Enlightenment scholarship used predominantly the national languages, for the benefit of anyone who wanted to educate themselves. She comes to the conclusion that the use of Latin as the language of scholarship was also disappearing. Ultimately, the gap between new terminology and Latinity, the struggle for clarity combined with beauty, also become too great in university education. All in all, the role of Latin as an academic passport became increasingly restricted." (Dykstra 2010, 239-240) "In the first half of the nineteenth century Latin lost its international prestige as a means of communication." (Dykstra 2010, 240)

[6] On Croatian Neo-Latin Literature see Birnbaum 1986, Vratović 1989, 1996; Gortan-Vratović 1960-1970, Jovanović-Luggin-Špoljarić-Šubarić 2018.

[7] For more about Croatian Church Slavonic see Mihaljević 2009, Mihaljević-Reinhart 2005, Gadžijeva et al. 2014.

[8] Gadžijeva et al. 2014, 18, Mihaljević 2011.

texts.[9] On the relation between Latin and Church Slavonic see e.g. Galabov 1973, Keipert 1987, 1996, Mihaljević 2020.

From the thirteenth century onward, vernaculars in Europe went through a process of standardization and began competing with Latin in different fields. For a language to be truly standardized, the existence of a dictionary is one of the main prerequisites. As we know, dictionaries helped to stabilize linguistic norms and the existence of a dictionary was one of the main characteristics of a standard language. The importance of an authoritative dictionary has often been mentioned when discussing the definition of a standard language, e.g. "Typical characteristics of a standard language include some of the following: An authoritative dictionary which records the vocabulary of the language [and] An authoritative grammar which records the forms, rules and structures of the language."[10]

3. Dictionaries as a testimony of the linguistic world

The first dictionaries produced in Croatia or with Croatian as one of their component languages were not generally monolingual, but mostly multilingual with Latin as one of the languages included, differing from the development of dictionaries in general in which the first dictionaries to appear were monolingual.[11]

Polyglot or multilingual publications are considered a useful testimony of the linguistic world of a certain period as they offer linguistic,

[9] Gadžijeva et al. 2014

[10] http://www.ucl.ac.uk/dutchstudies/an/SP_LINKS_UCL_POPUP/SPs_english/language_history/pages/stand_lang.html

[11] "Boisson (1996) and Boisson et al. (1991) describe in great detail the history of bilingual lexicography. Contrary to what is frequently assumed, bilingual dictionaries came later than monolingual dictionaries and the lexicographic traditions in Mesopotamia and in Ancient Egypt were clearly originally monolingual because the great civilizations in the Middle East were, as is pointed out by Van Sterkenburg (2003b: 8-9), self-centered and did not focus on their neighboring countries." (Fontenelle 2015, 44)

social, and cultural data.[12] Especially interesting are multilingual publications for linguistic instruction such as dictionaries or grammars.[13]

Šime Demo (2011) gives data on Latin and Italian domination in Croatian dictionaries up to the eighteenth century[14], stating that a dictionary is a rare example in which these two languages continuously interact[15] and that it is an ideal opportunity for determining the relation between Latin and Croatian.[16] Demo stresses:

> "By a set of selected features (the fact that they have been written and published, the choice of the metalanguage and of the languages described, direct remarks of the author, methods of dealing with certain topics) dictionaries and grammar books of a certain language can show what cultural status that language had among the intellectual and political elite. The decline of the importance of Latin in Croatian culture, or denying its auxiliary role in sowing the seeds of the standard national language, can be very clearly traced in language manuals." (Demo 2011, 60)

Latin was, for the most part, the dominant language in the lexicography of other European countries as well until the seventeenth century.[17]

---

[12] "A particular category of sources, namely "polyglot publications"—used in this context as the overarching name for all books that contain more than one language, regardless of their organization, structure, or particular use—serve as an important testimony of the linguistic world of early modern Europe and offer vital information for the study of the social and cultural histories of various languages involved." (Verbeke 2015, 29)

[13] "An intentional and obvious form of polyglossia involving Latin and one or more vernaculars is found in publications used for linguistic instruction. This category contains, in addition to dictionaries and grammars, also phrasebooks and collections of dialogues or other texts enabling the user to compare the version of the language he or she knows with the version of the language he or she is trying to learn." (Verbeke 2015, 29)

[14] "... u rječnicima dominacija latinskoga i talijanskoga do polovice 18. st. nakon toga njemački do 19. st. od 20. st. engleski..." (Demo 2011, 53) ('…in dictionaries, Latin and Italian are the dominant languages until the mid-eighteenth century, after that German is the dominant language until the 19th and English until the 20th centuries…').

[15] "... jezični su priručnici rijetka vrsta djela u kojima se dva jezika eksplicitno i kontinuirano dovode u međuodnos." (Demo 2011, 52) ('… linguistic manuals are a rare type of books where two languages are in explicit and continuous interaction').

[16] "... idealno za proučavanje odnosa latinskoga i hrvatskoga…" (Demo 2011, 52) ('… ideal for the study of the relation between Latin and Croatian…').

[17] "Latin continued to dominate bilingual lexicography right into the seventeenth century, and also formed the source language for polyglot vocabularies such as the

4. Latin and Italian influence in dictionaries and word lists before Vrančić

Vrančić's *Dictionarium quinque nobilissimarum Europae linguarum* published in Venice in 1595 is often considered to be the first dictionary in the history of Croatian language and literacy.[18] Even before Vrančić, however, there were a few shorter dictionaries or word lists which included Croatian as one of the languages. Following the vernacular glosses added to Latin works,[19] which mark the beginning of lexicography in most European countries, the first word lists and short dictionaries appear. In Croatia, these glosses can be found in Croatian Church Slavonic texts as well as in those written in Croatian, Latin, and Italian. In those texts, glosses can be written using the Latin or Italian equivalents of the original word, including synonyms, the same word written in a different script, or an explanation of the meaning. The appearance of the gloss can be a sign that the original word is an archaism.[20] Some Kajkavian glosses appear in the Radon Bible.[21]

The first text with a word list including a Croatian column was the *Liber de simplicibus*[22] by Nicolò Roccabonella (1386-1459), an Italian doctor from Conegliano, who after working for three decades in Venice, lived in Zadar from 1449 to 1453. He wrote his *Liber* as a practical tool for trade with the goal of avoiding misunderstandings. Thus, he added pictures to his word list. The *Liber* consisted of 370 Slavic plant names with 330 Croatian words. From the example of Roccabonella's dictionary, it is obvious that the first dictionary came about as the result of Italian influence and Croatia's proximity to Italy. Pharmaceutics began developing in Italy and southern France and spread to Dalmatia due to Venetian influence. After Roccabonella's *Liber,* two foreign dictionaries with Croatian words also show that the first dictionaries were written by foreigners and

---

popular *Sex linguarum* and Calepino's monumental *Dictionarium* (see Stein 1985), even though these contained perfectly familiar and concrete everyday words." (Snell-Hornby 1986, 213)

[18] Vrančić. 1971. A digitalized version of the dictionary is available on the PHRB using the following link: http://crodip.ffzg.hr/. Dukat 1925, Jonke 1971, Franolić 1976, Vončina 1979, Samardžija 2005.

[19] "The first explained Latin words in a vernacular language, but they were soon reversed to produce dictionaries explaining the words of the vernacular language in Latin." (Béjoint 2015, 9-10)

[20] Marti 1987.

[21] Filipović 1986, 65.

[22] Gostl 1955, Vajs 1995.

influenced by their native language. In his *Pilgrimage,*[23] Arnold von Harff (1471-1505) described his travel through Croatia. He added 56 Croatian words to his description. He wrote Croatian words as he heard them. Here are some examples of the words from his word list, with the standard Croatian word and the English translation given in parentheses: *benese* (penezi = novac, 'money'), *besenitz* (pšenica, 'wheat'), *bytte* (piti, 'to drink'), *crochga* (G kruha, 'bread.GEN.SG'), *dreuo* (drvo, 'wood/tree'). The Croatian column appears in the work of the Italian author Pietro Lupis Valentian entitled *Opera nuova che insegna a parlare a lingua schiavonesca alli grandi, alli picoli et alle donne*, published in Ancona in 1527.[24] The book consists of an Italian-Croatian dictionary with 412 Italian and 328 Croatian words. The Croatian words were written using Italian orthography. Here are some examples: *cerqua* (crkva, 'church'), *chiachia* (ćaća, 'father'), *cocoxe* (kokoše, 'chicken'), *cobluc* (klobuk, 'hat'), *cosuglia* (košulja, 'shirt'), *coxa* (koža, 'leather/skin'), *cozuchar* (kožuhar, 'seller/maker of sheepskin coats'), *crusca* (kruška, 'pear'), *cruch* (kruh, 'bread'), *deuoica* (djevojka, 'girl'), *iaiza* (jajca, 'eggs').

Some Croatian Church Slavonic words were attested in Sigismund Gelenius' (1497-1554) dictionary of similarly sounding words (paronyms, false friends, equivalents) *Lexicon symphonum quo quatuor linguarum Europae ... concordia ... indicatur*[25] published in Basel in 1537. It is a dictionary of Latin, German, and Greek which cites some Slavonic words. Among them, there are some Croatian Church Slavonic words from the Emmaus monastery. Sigismund Gelenius studied in Italy; therefore, he was also under the influence of Italian culture and language.

Croatian Neo-Latin authors were among the first Croats to have an impact on the development of Croatian dictionaries. Bartol Đurđević / Bartholomaeo Gyurgieuits / Bartholomaeo Georgieuiz / Bartholomeo Georgieuitz de Croacia, detto Pellegrino Hierosolymitano (1506-1566) created a word list consisting of a number of Croatian words and their Latin translations, as well as some phrases in both languages. His work *De aflictione tam captivorum quam etiam sub Turcae tributo viventium christianorum*[26] published in 1544 in Anvers was translated into many languages, with the translators also translating the Latin part of the dictionary into their native language. Thus, the fact that Croatian Neo-Latin authors were read throughout Europe was also a vehicle for the spread of words from the vernacular.

[23] Letts 2010.
[24] Putanec 1979.
[25] Gelenius 1573.
[26] Đurđević 1545.

Translation was also a stimulus for dictionary compilation. Two translators who translated from Latin added explanations of their translation equivalents. Especially interesting is Ivan Pergošić's translation of the work *Tripartit* (*Opus tripartitum iuris consuetudinarii inclyti Regni Hungariae partiumque eidem annexarum*) because he translated legal texts and created the first terminological dictionary *Decretum koteroga ie Verbewczi Istvan diachki popiszal*. *Ve teh szu knighah ve te diachke ryechi ouako pretumachene*, including 45 Latin terms.

Two Slovene dictionaries also contain some Croatian words: Šime Budinić – *Pochorni i mnozii ini psalmi Davidovi, sloxeni v slovignschi iazich na cislo, i miru: po Scymunu Budineii popu Zadraninu* published in Rome in 1582—*Otvorenye, il tumacenye nichogijh rijcij, poloxenih u ouih psalmeh: choye nijsu usachomu opcchiene ni occitte*[27]—123 words with synonyms; and the dictionary by Juraj Dalmatin as an addition to the work *Biblia, tu ie, Vse Svetu Pismu, Stariga inu Noviga Testamenta, slovenski, tolmazhena...*[28] published in Ljubljana in 1578. Croatian equivalents are also given in Hieronymus Megiser's *Dictionarium quatuor linguarum*, *videlicet Germanicae, Latinae, Illyricae (quae vulgo Sclavonica appelatur) et Italicae seu Hetruscae*[29] published in 1592 in Graz.

5. Latin and Italian influence in Vrančić's dictionary

The first comprehensive Croatian dictionary is Faust Vrančić's *Dictionarium quinque nobilissimarum Europae linguarum – Latinae, Italicae, Germanicae, Dalmatiae, & Ungaricae*[30] published in 1595 in Venice. Although it has Latin as the entry language and the entry language is mostly considered to be the dominant language, it is important for the history of Croatian lexicography that one of the languages is Croatian (called *Dalmatian* in the dictionary).

Vrančić was educated both in Venice and Padua, and that is where he got the idea to write the dictionary. Vrančić's dictionary is considered to be one of the first Hungarian dictionaries as well. It also offers example illustrating how ideas flowed from Italy to Croatia and then to Hungary and other European countries. It has been hypothesized that the author used Calepinus' dictionary *Septem linguarum Calepinus, hoc est, Lexicon Latinum* and Estienne's *Dictionarium, seu Latinae linguae thesaurus* for

---

[27] Budinić 1582, Putanec 1982.
[28] Dalmatin 1578.
[29] Megiser 1592.
[30] Vrančić 1971.

selecting Latin dictionary entries although the word list is not the same[31]. Latin and Italian influenced Croatian equivalents. Often, Latin loanwords are given as Croatian equivalents such as *aër* ('air'). There are also many multiword units as translation equivalents: *oleaster – divyaamaßlino* ('wild olive'), *palumbis – divvyi-golub* ('wild pigeon'), *caprea – divya koza* ('wild goat'), *labrusca – divya-loza* ('wild grape'), *caprificus – divya szmokva* ('wild fig'), and definitions *recuperare – dobiti-oppeta* ('to regain'), *largiri – dati-obilno* ('to lavish, bestow'), *animare – davati-ßarcze* ('to animate, inspire').

Many noun phrases have the following structure *noun* + *preposition* od ('for') + *noun.GEN*: *natatis – dă-od-rodyenya* ('birthday'), *carina – dno-od-plavi* ('keel, bottom of ship'), *armamentarium – mißzto-od-orusya* ('armoury'). This construction is generally considered to have arisen due to Italian influence, but in this specific case, the Italian equivalents in this dictionary do not use that construction: *natale, carina, arsenale*. In this case, the noun phrase was likely influenced by spoken Italian. Vrančić wrote multiword equivalents either as a single word *oleaster – divyaamaßlino* ('wild olive'), or more often with hyphens *fere, ferme, pene – malo-dane* ('almost'), *quoquouersum – na-ßve-ßtranè* ('in every direction'), *quamvis – budi-da* ('although'). Rarely, they are written as multiword expressions: *quadrans – csetvarti dil* ('quarter, forth part'), *caprea – divya koza* ('wild goat'). There is one example of a two-word expression separated by a comma, perhaps as the result of a typographical error: *quamprimum – csa,barxè* ('as soon as possible'). In his dictionary some noncanonical words appear, such as forms of the verb *to be*: *esto – budi* (to be.2.SG.IMP.) or of the pronoun *I*: *mihi – meni* (I.DAT). In some cases, Croatian participles appear as equivalents of Latin adjectives which were originally participles: *vigilans – bdechij* ('watchful, vigilant').

Vrančić's dictionary influenced Bernardino Baldi, who wrote a Hungarian-Italian dictionary although he presumably did not speak Hungarian, and Megiser's dictionary *Thesaurus polyglottus vel Dictionarium multilingue* written in 1603.[32]

The most important dictionary to have been written under the influence of Vrančić, however, is the dictionary by Petar Loderecker (Petrus Lodereckerus Pragenus) entitled *Dictionarium septem diversarum linguarum, videlicet Latine, Italice, Dalmatice, Bohemice, Polonice, Germanice et Ungarice*[33] and published in 1605 in Prague[34]; he added

[31] Birnbaum 2001, Stankiewicz, 1984.
[32] Bago 2014: 26.
[33] Loderecker 2005.
[34] Samardžija 2005, Katičić 2005.

Czech and Polish equivalents to Vrančić's dictionary, and he also compiled the first dictionary with Croatian as the entry language. It is a typical example of how dictionaries were compiled at the time—the authors would change column order from the existing dictionaries. The purpose of such dictionaries was to teach students foreign languages. That is how multiword units and definitions become dictionary entries. This is also an example of intensive contact between Czech and Croatian.

6. Bartol Kašić's *Croatian–Italian dictionary*

The handwritten and unpublished Croatian-Italian dictionary of Bartol Kašić (1575-1650) has a similar dictionary structure.[35] Marin Temparica, a sixteenth-century Jesuit from Dubrovnik, explained the reasons behind Kašić's decision to compile a dictionary—he wanted to help students learn the language: *apresso fare un vocabulario, e una gramatica per aiutare li studenti (ancor forastieri*)[36]. In his autobiography Kašić gave a similar explanation: *ad addiscendam facilius linguam instituerentur*[37]. Kašić compiled a dictionary for students and foreigners to learn the language. Therefore, Croatian-Italian bonds influenced the compilation of this dictionary as well. The dictionary was written for students traveling to Italy. The author also used Vrančić's dictionary. Kašić's dictionary consists of 3700 Croatian words, 3120 of which are included in Vrančić's dictionary, which has 3581 words in total. In his dictionary, Kašić changed the order of the columns. Under Vrančić's influence, however, some uncanonical forms of words appear as dictionary entries, and some multiword units are written as one word. The relation between these two dictionaries is illustrated using selected examples in Table 13-1. Kašić mostly omitted multiword entries or wrote them as a single word. Participles and noncanonical forms are included in the dictionary.

[35] Horvat 1999. A digitalized version of the dictionary is available on PHRB at the following link: http://crodip.ffzg.hr/.
[36] Horvat 1999. ('then write a dictionary and a grammar to help students (still foreigners)')
[37] *ad addiscendam facilius linguam instituerentur* 'is compiled for easier language learning'

Vrančić	**Kašić**
bdechij – vigilans ('vigilant')	*bdeći*
bixechi – fugax ('avoiding')	*bježeći*
csa,barxè – quamprimum ('as soon as possible')	*ča brže*
on-csaß – subito ('at once')	*ončas*
csetvarti dil – quadrans ('quarter, forth part')	
sztraxu csiniti – excubare ('lie in the open')	
vecserße-csini – advesperascit ('evening is coming on')	
dă-od-rodyenya – natatis ('birthday')	
drugi-dan – postridie ('on the following day')	
nikidan – nuper ('recently')	*nikidan*
dva dni – biduum ('two days')	
svaki dan – quotidie ('daily')	
krivinu davati – culpare ('to blame')	

Table 13-1: The differences between Vrančić's and Kašić's dictionaries.

7. Jakov Mikalja's *Thesaurus Linguae Illyricae*

The first modern dictionary in Croatian linguistic history is Jakov Mikalja's (born Peschici, 1601 – died Loreto, 1654) dictionary, *Thesaurus Linguae Illyricae – Blago jezika slovinskoga*[38] published in 1649/1651.[39] It is a trilingual dictionary—the entry language is Croatian, and the other two languages are Italian and Latin. It is composed of 928 pages with approximately 25,000 entries. Once again, the author states that the purpose of the dictionary is to help priests in the Loreto Illyric collegium learn Croatian prior to their mission and to help students learn Italian and understand religious books. The dictionary entries were probably taken from a Latin dictionary[40]. I will show, however, that the author likely predominantly used an Italian dictionary. There are many loan translations from Italian and neologisms and loan words from Italian and Latin. In many entries, the translation equivalents are definitions of Italian or Latin words: cro. *izniti zao glas na koga* (bring out.INF evil.ADJ.ACC.SG voice.ACC.SG on someone.INDEF.ACC 'to defame someone') – tal.

[38] Mikalja 2011.
[39] The dictionary has been transcribed and interpreted in Gabrić-Bagarić et al. 2011.
[40] Gabrić-Bagarić et al. 2011, 953.

infamare alcuno (defame.INF someone) – lat. *infamo* (defame.1.SG.PRS). From this example, it is not evident whether this word was translated from Italian or Latin. However, if we analyze other multiword units, it is clear that Croatian equivalents are often a literal translation from Italian while literal translations from Latin are sometimes used as an explanation. In Table 13-2 some examples are given in which the Italian and Latin columns differ, and the Croatian column follows the Italian structure.

Croatian	**Italian**	**Latin**
davati krivinu give.INF blame.ACC.SG 'to blame someone' (lit. 'to give someone the blame')	*dar la colpa* give.INF blame 'to blame someone' (lit. 'to give someone the blame')	*neglegentiae expensum ferre* negligence.GEN.SG weight.ACC.SG bear.INF 'to blame someone' (lit. 'to carry the weight of negligence')
ostaviti za mrtva (izbiti na smrt) leave.INF for dead.ADJ.ACC.SG. 'to beat to death' (lit. 'to leave for the dead')	*lacsjar per morto* leave.INF for dead.ADJ 'to beat to death' (lit. 'to leave for the dead')	*deuerberare usque ad mortem* beat.INF all the way.ADV to death.ACC.SG 'to beat to death'
otiti s dobrom srićom leave.INF with good.ADJ.INS.SG luck.INS.SG. 'to leave with good luck'	*andarsene con buona ventura* leave.INF with good.ADJ fortune 'to leave with good luck'	*abire in bonam rem* depart/go off.INF into good.ADJ.ACC.SG thing.ACC.SG 'to leave with good luck' (lit. 'to leave in good thing')
imati veliku pamet have.INF a great.ADJ.ACC.SG mind.ACC.SG 'to be smart' (lit. 'to have a great mind')	*hauer grand' ingegno* have.INF a great.ADJ mind 'to be smart' (lit. 'to have a great mind')	*praestare, excellere, florere, valere, pollere ingenio, esse acutissima ingenii acie* be outstanding.INF, excel.INF, flourish.INF, be powerful.INF, be strong.INF brain.ABL.SG, be.INF the_most_sharp/wise.ADJ.ABL.SG brain.GEN.SG sharpness.ABL.SG 'to be outstanding, to excel, to flourish, to be powerful, to be smart (lit. 'to be of the sharpest brain')'

pjenezi dati na dobit money.NOM.PL. give.PTCP.PRF.PAS.NOM.PL to profit.ACC.SG ‘money given for profit’	*denari dati a guadagno* money give.PTCP.PRF.PAS. to profit ‘money given for profit’	*pecunia quaestuosa* money.NOM.SG profitable.ADJ.NOM.SG ‘money given for profit’ (lit. ‘profitable money’)
promisliti dobro jednu stvar think.INF good.ADV one.ART.ACC.SG thing.ACC.SG ‘to consider a thing in-depth’ (lit. ‘to think well about a thing’)	*considerar bene una cosa* consider.INF well.ADV one.ART thing ‘to consider a thing in-depth’ (lit. ‘to think well about a thing’)	*perpendo pensum, pondero, recte examino…* weigh carefully.1.SG.PRS task.ACC.SG, weigh.1.SG.PRS, properly.ADV examine.1.SG.PRS ‘to consider a thing in-depth’ (lit. ‘to weigh carefully’)
staviti ruku na koji posao put.INF hand.ACC.SG on someART.ACC.SG job.ACC.SG ‘to undertake a job’ (lit. ‘to put a hand on the job’)	*metter la mano in qualche negotio* put.INF hand on some job ‘to undertake a job’ (lit. ‘to put a hand on the job’)	*suscipere negotium* undertake.INF job.ACC.SG ‘to undertake a job’
staviti se u nevolju place.INF REFL.ACC in trouble.ACC.SG ‘to get into trouble’ (lit. 'to place oneself into trouble’)	*mettersi in necessita* place.INF.REFL in need ‘to get into trouble’ (lit. ‘to place oneself into trouble’)	*imponere sibi necessitatem* place.INF REFL.DAT necessity.ACC.SG ‘to get into trouble’ (lit. ‘to give trouble to oneself’)
činiti se t(i)jelo make.INF REFL.ACC body.NOM.SG ‘to incorporate’ (lit. ‘to make oneself a body’)	*farsi corpo, incorporarsi* make.INF.REFL. body, incorporate.INF.REFL ‘to incorporate’ (lit. ‘to make oneself a body’)	*corporor, in corpus coalescere* form into body.1.SG.PRS.PAS, in the body.ACC.SG become unified.INF ‘to incorporate, to grow/join together’
trgati život (mučiti se) tear.INF life.ACC.SG (agonize.INF.REFL. ACC)	*stentar la uita* struggle.INF life ‘to torture oneself, to agonize’	*duriter vitam agere* hard.ADV life.ACC.SG conduct.INF ‘live a hard life’

‘to torture oneself, to agonize’ (lit. 'to tear the life’)	(lit. 'to tear the life’)	
voditi za nos lead.INF by the nose.ACC.SG ‘to deceive’ (lit. ‘to lead by the nose’)	*menar per il naso* run.INF by the nose ‘to deceive’ (lit. ‘to run by the nose’)	*dare verba alicui* give.INF words.ACC.PL to someone.INDEF.DAT. ‘to deceive, to give empty words’ (lit. ‘to give words to someone’)
poći od srca [isprazniti crijeva] go.INF from the heart.GEN.SG [empty.INF the bowels.ACC.PL] ‘to defecate’ (lit. 'to go from the heart’)	*andar del corpo* go.INF from the body ‘to defecate’ (lit. 'to go from the body)	*caco, exonerare ventrem, excerno, egero* defecate.1.SG.PRS, unload.INF stomach.ACC.SG, discharge.1.SG.PRS, carry out.1.SG.PRS ‘to defecate’ (lit. ‘to defecate, to unload stomach’)

Table 13-2: The differences between Croatian, Italian, and Latin columns.

Many hapax legomena and words that are attested for the first time appear in this dictionary. For some of them it is impossible to say if they were borrowed from Italian or Latin because the Italian word is derived from the Latin one, for example *akolitat* (‘acolyte, cleric in minor orders’) – *acolitato* – *acolitatus*. However, in some examples it is again evident that the Croatian column is formed according to the Italian column. In Table 13-3, examples which show that Croatian words were formed according to the Italian column are given, while the Latin column differs considerably.

Croatian	Italian	Latin
arčipelag ‘archipelago’	*l’arcipelago* ‘the archipelago’	*mare Aegeum* sea.NOM.SG. Aegean.ADJ.NOM.SG ‘Aegean Sea’
armiralja ‘admiral’	*armiraglja* ‘admiral’	*praefectus orae* commander.NOM.SG of the shore.GEN.SG ‘admiral’
arsenik ‘arsenic’	*arsenico* ‘arsenic’	*auripigentum maritimae* orpiment.NOM.SG. maritime/of the sea.ADJ.GEN.SG ‘arsenic’ (lit. ‘the maritime orpiment’)
busula ‘compass’	*bossola* ‘compass’	*acus nautica versoria* needle.NOM.SG. nautical.ADJ.NOM.SG. that guides the ship.ADJ.NOM.SG ‘compass’ (lit. ‘nautical needle that guides the ship’)
catara ‘raft’	*zattera* ‘raft’	*rates* ‘raft’
gamura ‘female's light outer garment with decorative border’	*gamorra* ‘female's light outer garment with decorative border’	*cyclas* ‘female's light outer garment with decorative border’
gulaf ‘gulf’	*golfo* ‘gulf’	*sinus* ‘bay’
inčensir ‘censer’	*incensiero* ‘censer’	*thurribulum* ‘censer’

Table 13-3: Croatian hapaxes following the Italian model.

The Italian influence is visible on the syntactic level as well. Due to the influence of Italian, many noun phrases contain the preposition *od* (‘for’) as a translation of Italian *di* and *da*, as in Vrančić’s dictionary. These examples are given in Table 13-4. In Croatian language history, a similar influence was evident while translating from Latin into Croatian and Croatian Church Slavonic, but the scope of this phenomenon is also probably a result of the influence of spoken Italian.[41]

[41] Mihaljević 2020.

Croatian	**Italian**	**Latin**
daž od krvi rain.NOM.SG. of blood.GEN.SG 'rain of blood, bloody rain' (lit. 'rain of blood')	*pjoggja di sangue* rain of blood 'rain of blood, bloody rain' (lit. 'rain of blood')	*imber sanguineus* rain.NOM.SG bloody.ADJ.NOM.SG 'rain of blood, bloody rain' (lit. 'bloody rain')
djevojka od udadbe girl.NOM.SG of marriage.GEN.SG 'girl fit for marriage' (lit. 'girl of marriage')	*vergine da marito* virgin/girl of a husband 'girl fit for marriage' (lit. 'girl of a husband')	*virgo nubilis, adulta* maiden/virgin/girl.NOM.SG marriageable.ADJ.NOM.SG, adult.ADJ.NOM.SG 'girl fit for marriage' (lit. 'marriageable girl, adult girl')
drug od skule friend.NOM.SG from school.GEN.SG 'school friend' (lit. 'friend from school')	*compagnio di scuola* 'school friend' (lit. 'friend from school')	*condiscipulus* fellow pupil.NOM.SG 'school friend' (lit. 'fellow pupil')
kamenica od svete vode aspersorium.NOM.SG of holy.ADJ.GEN.SG water.GEN.SG 'aspersorium' (lit. 'aspersorium of holy water')	*pila dell' acqua santa* aspersorium of water holy.ADJ 'aspersorium' (lit. 'aspersorium of holy water')	*pila aquae benedictae* aspersorium.NOM.SG of water.GEN.SG blessed/holy.ADJ.GEN.SG 'aspersorium' (lit. 'aspersorium of holy water')
otrnutje od zubi numbness.NOM.SG of teeth.GEN.PL 'numbness of teeth'	*gleamento di denti, legamento di denti* gleaming of teeth, ligament of teeth 'numbness of teeth'	*stupor dentium* numbness.NOM.SG of teeth.GEN.PL 'numbness of teeth'
riba od rike fish.NOM.SG from the river.GEN.SG 'freshwater fish' (lit. 'fish from the river')	*pesce di fiume* fish from the river 'freshwater fish' (lit. 'fish from the river')	*piscis fluvialis vel fluviaticus* fish.NOM.SG fluvial.ADJ.NOM.SG or from the river.ADJ.NOM.SG 'freshwater fish' (lit. 'fluvial fish')

rođen od kraljeve krvi born.PTCP.PRF.PAS.NOM.SG of king's.ADJ.GEN.SG blood.GEN.SG 'born of royal blood'	*nato di sangue reale* born.PTCP.PRF.PA of the blood royal.ADJ 'born of royal blood	*stirpe regia ortus; statu regio editus* lineage.ABL.SG royal.ADJ.ABL.SG born.PTPC.PRF.PAS.NOM.SG, status/position.ABL.SG royal.ADJ.ABL.SG elevate/rise.PTPC.PRF.PAS.NOM.SG 'born of royal blood'

Table 13-4: Nominative + *od* + genitive structure.

In some examples, the Croatian word is not a literal translation of Italian components, but the syntactical structure is the same: *glava od kuće* (head.NOM.SG of house.GEN.SG 'head of the house') – *padre di famiglja* ('father of the family') – *pater familias* ('father of the family'); *trenutje od oka* (moment.NOM.SG of the eye.GEN.SG 'moment of the eye') – *un batter d'occhi* (blink of eye.PL 'a blink of the eyes') – *oculi momentum* (eye.GEN.SG moment.NOM.SG 'moment of the eye'), *ictus oculi* (blink.NOM.SG eye.GEN.SG 'a blink of an eye').

Italian influence is seen in the usage of the construction *za* ('for') with the infinitive denoting purpose, which is a direct translation of Italian *per + infinitive*. Examples of this construction are given in Table 13-5. The construction appears even in examples in which the Italian equivalent does not have that structure.

Croatian	Italian	Latin
za proći vr(i)jeme for pass.INF time.NOM.SG 'for recreation' (lit. 'for the time to pass')	*per recreatione* 'for recreation'	*animi causa* soul/heart.GEN.SG for 'for recreation' (lit. 'for the soul/heart')
molba i nastojan'je za imati koju čast plea.NOM.SG and effort.NOM.SG to have.INF some.INDEF.ACC.SG honor.ACC.SG 'plea and effort to gain an honor'	*brogljo* 'fraud'	*praensatio* fraud.NOM.SG 'fraud'

moliti mnogo i nastojati za imati koju čast pray.INF a lot.ADV and strive.INF to have.INF some.INDEF.ACC.SG honor.ACC.SG 'to plea a lot and try to gain an honor'	*far brogljo* make.INF fraud 'to fraud' (lit. 'to make fraud')	*praenso* fraud.1.SG.PRS 'to fraud'
staviti ruku na krv, za učiniti krv put.INF hand.ACC.SG on blood.ACC.SG, to make.INF blood.ACC.SG 'to plan murder' (lit. 'to put a hand on blood, to make blood')	*metter mano in sangue* put.INF hand in blood 'to plan murder' (lit. 'to put a hand on blood')	*spectare ad caedem* consider/look.INF at murder.ACC.SG 'to plan murder' (lit. 'to look on murder')

Table 13-5: *Za* + infinitive structure.

While translating from Italian and Latin into Croatian or Croatian Church Slavonic in the medieval and early modern period, translators often copied the word order of the original.[42] At first glance, one would be under the impression that such is the case in the dictionary entries as well, but if we look at the examples more closely, it becomes clear that word order does not reflect the Italian word order, depending instead on the entry word. If the entry word is a noun, the noun will come first, but if the entry word is an adjective, the adjective will come first. Examples of this are given in Table 13-6.

Croatian	**Italian**	**Latin**
c(i)jena učinjena s. v. c(i)jena price.NOM.SG do.PTCP.PRF.PAS .NOM.SG – price.NOM.SG 'price agreed upon' (lit. 'price done')	*prezzo patuito* price agree.PTCP.PRF.PA S.NOM.SG 'price agreed upon'	*pretium pactum* price.NOM.SG agree.PTCP.PRF.PAS.NOM.S G 'price agreed upon'
*glavna stvar s. v. *glavni* main.ADJ.NOM.S G thing.NOM.SG – main.ADJ.NOM.SG 'main thing'	*somma d'vna cosa* the sum of one thing 'essence of a thing'	*caput rei* head.NOM.SG thing.GEN.SG 'essence of a thing'

[42] Mihaljević 2020.

gluho doba noći s. v. gluh deaf.ADJ.NOM.SG time.NOM.SG night.GEN.SG – deaf.ADJ.NOM.SG 'dead of night'	*il tempo della notte che non si sente un zitto* the time of the night which no hear.3.SG. REFL. a shut up 'dead of night' (lit. 'the time of the night when you cannot hear anything')	*conticinium* quiet of the night/dead of night.NOM.SG 'dead of night'
*kozja brada s. v. *kozji* goat's.ADJ.NOM.SG beard.NOM.SG – goat's.ADJ.NOM.SG 'goat's beard'	*barba di capro* beard of goat 'goat's beard' (lit. 'beard of goat')	*aruncus* goat's beard.NOM.SG 'goat's beard'
mir pravi s. v. mir peace.NOM.SG real.ADJ.NOM.SG – peace.NOM.SG 'real peace'	*pace certa* peace certain.ADJ 'real peace' lit. 'certain peace')	*pax explorata* peace.NOM.SG explore/investigate.PTCP.PRF.PAS.NOM.SG 'real peace' (lit. 'explored peace')
smrtna nemoć s. v. smrtni mortal.ADJ.NOM.SG infirmity.NOM.SG – mortal.ADJ.NOM.SG 'mortal infirmity'	*infirmita mortale* infirmity mortal.ADJ 'mortal infirmity'	*infirmitas mortifera* infirmity.NOM.SG deadly/death-bringing.ADJ.NOM.SG 'mortal infirmity'
dragi kami s. v. jačintić precious.ADJ.NOM.SG stone.NOM.SG – sapphire.NOM.SG 'precious stone, gemstone'	*pietra pretiosa* stone precious.ADJ. 'precious stone, gemstone'	*Hyacinthus* sapphire.NOM.SG 'sapphire'
u kuću moju s. v. kuća in the house.ACC.SG my.POSS.ACC.SG – house.NOM.SG 'in my house'	*in casa mia* in the house my. POSS 'in my house'	*apud me* at I.PERS.ACC 'in my house'

Table 13-6: Noun and adjective word order.

The adjective comes before the noun if neither adjective nor noun is the headword—regardless of the Italian or Latin word order. The situation is similar with verbs and adverbs or adverbs and adjectives: *sin posinjen s. v. posiniti* (son.NOM.SG adopte.PTCP.PRF.NOM.SG – adopte.INF, 'adopted son') – *figlio adottiuo* (son adopted.ADJ) – *filius adoptatus* (son adopte.PTCP.PRF.NOM.SG). Exceptions to this rule usually follow the Italian word order, as shown in Table 13-7.

Croatian	**Italian**	**Latin**
l(i)jepa prigoda s. v. prigoda beautiful.ADJ.NOM.SG occasion.NOM.SG – occasion.NOM.SG 'nice occasion'	*bella occasione* beautiful.ADJ occasion 'nice occasion'	*praecllara, amplia, mirifica occasio* splendid.ADJ.NOM.SG, great.ADJ.NOM.SG, amazing.ADJ.NOM.SG, wonderful.ADJ.NOM.SG opportunity.NOM.SG 'splendid, nice, great, amazing, wonderful occasion'
mala spomena / mala pamet s. v. spomena small.ADJ.NOM.SG memory.NOM.SG / small.ADJ.NOM.SG mind.NOM.SG – memory.NOM.SG 'small souvenir, short memory'	*poca memoria* little.ADJ memory 'small souvenir, short memory'	*memoriola* memory.NOM.SG.DEM 'small souvenir, short memory'
stvar suđena s. v. suditi thing.NOM.SG judge.PTCP.PRF.PAS.NOM.SG – judge.INF 'sentence, judgment' (lit. 'judged thing')	*cosa gjudicata* thing judge.PTCP.PRF.PAS.SG 'sentence, judgment' (lit. 'judged thing')	*judicatum* sentence/judgment.PTCP.PRF.PAS.NOM.SG 'sentence, judgment' (lit. 'judged thing')
izgubiti ljubav i milost stečenu s. v. milost lose.INF love.ACC.SG and grace.ACC.SG acquire.PTCP.PRF.PASS.ACC.SG – grace.NOM.SG 'lose gained love and grace'	*perder la gratia acquistata* lose.INF the grace acquire.PTCP.PRF.PASS.SG 'lose gained grace'	*effundere gratias collectas* lose.INF the favors.ACC.PL obtaine.PTCP.PRF.PAS.ACC.PL 'lose gained grace'

Table 13-7: Exceptions from the general rules.

Adjective *zao* ('mean') almost always comes first regardless of the entry word*: zao jezik* ('evil tongue') – *lingua mordace* ('biting tongue') – *acida lingua* ('bitter tongue'), *zla ćud* ('evil temper') – *mala indole* ('bad temper') – *mali mores* ('evil customs').

8. After Mikalja

After Mikalja there were two other important dictionaries with the Latin column written prior to the eighteenth century. The entry words in these dictionaries are similar to Mikalja's, and they show a similar Latin and Italian influence, although to a lesser extent. Following these dictionaries, Latin became less important in Croatian lexicography while the influence of German grew. In modern Latin-Croatian dictionaries, Latin influence is visible only in a small number of loan words and in the word order of some examples.

9. Conclusion

The influence of Italian and Latin in Croatian lexicography is evident on different levels. The idea of compiling a dictionary that has a Croatian column either came from Italian authors or from authors who studied or lived in Italy, and the dictionaries were mainly geared toward Italian missionaries coming to Croatia or Croatian students going to study in Italy. Linguistic ideas spread from Italy to Croatia and then to other European countries—mostly the Kingdom of Bohemia, Hungary, and Poland. Italian influenced the orthography of some dictionaries. Influence can also be seen in the choice of entry words. It has been hypothesized that most Croatian dictionaries were modeled according to Latin and/or Italian dictionaries.[43] These languages had a strong influence on the grammatical structure of Croatian. The entries were written under an Italian influence—there are

---

[43] Most early modern dictionaries were modeled according to Latin ones: "Instead, scholars, translators, travelers, and diplomats alike were constrained either to speak Latin or to rely on vernacular glosses appended to a rather inferior Latin dictionary. This was the *Dictionarium* of Ambrogio Calepino, an Augustinian friar living in Bergamo. Calepino's original edition (1502) was a Latin vocabulary, with glosses in Latin supported by citations, together with encyclopaedic entries for the figures of classical mythology. In a second edition, glosses in Italian and French were added. By a process of accretion, the vocabularies of other languages, starting with Greek and Hebrew, were gradually added by others to successive editions of Calepino's original. In the words of Freed (2007), 'it evolved into the first polyglot dictionary'." (Hanks 2013, 12)

many loan translations and literal translations—so an influence on the grammatical structure can be detected. Dictionaries that have Croatian as the entry language (source language) were mostly compiled by reversing the columns of dictionaries which had another language as the entry language. Italian influence on Croatian is more evident than Latin, likely because it was a spoken language.

Bibliography

Bago, Petra. 2014. "Model višerazinske prezentacije tekstova starijih hrvatskih rječnika." PhD diss. University of Zagreb.

Béjoint, Henri. 2015. "Dictionaries for General Users: History and Development; Current Issues." In *The Oxford Handbook of Lexicography*, edited by Philip Durkin, 7-24. Oxford: Oxford University Press.

Birnbaum, Marianna D. 1986. *Humanists in a Shattered World: Croatian and Hungarian Latinity in the Sixteenth Century*. Columbus, Ohio: Slavica.

Birnbaum, Marianna D. 2001. "Ambrogio Calepino i Faust Vrančić." *Zbornik o Faustu Vrančiću*, edited by Vilijam Lakić, 31-35. Šibenik: Gradska knjižnica Juraj Šižgorić.

Budinić, Šime. 1582. "Pochorni i mnozii ini psalmi Davidovi, sloxeni v slovignschi iazich na cislo, i miru." Rome: Francesco Zanetti. Accessed January 14, 2020. http://dikaz.zkzd.hr/?vdoc=79&page=1

Considine, John. 2017. *Lexicography and Fieldwork in Post-Medieval Europe.* Oxford: Oxford University Press.

Dalmatin, Juraj. 1584. "Biblia, tu ie, Vse Svetu Pismu, Stariga inu Noviga Testamenta." Wittemberg: Hans Kraffts Erben. Accessed January 14, 2020. https://www.europeana.eu/hr/item/2020127/URN_NBN_SI_DOC_8NAJPUXS

Demo, Šime. 2011. "Latinski jezik kao nadnacionalni element u Hrvatskoj: primjer jezičnih priručnika." In *Etnicita Slovanského areálu: Historické proměny a současný stav,* edited by Kateřina Kedron and Marek Příhoda, 51–66. Praque: Červený Kostelec.

Derksen, Rick. 2019. "The Slavic and Baltic Languages." In *The Cambridge World History of Lexicography,* edited by John Considine, 445-459. Cambridge: Cambridge University Press.

Dukat, Vladoje. 1925. "Rječnik Fausta Vrančića." *Rad JAZU,* 231 (101): 102-136.

Dykstra, Anne. 2010. "Joost Halbertsma, Latin, and the Lexicon Frisicum" In *Adventuring in Dictionaries: New Studies in the History of Lexicography*, edited by John Considine, 238-255. Cambridge: Cambridge Scholars Publishing.

Đurđević, Bartul. 1545. "De afflictione tam captivorum quam etiam sub Turcae tributo viventium Christianorum." Accessed January 14, 2020. http://virtualna.nsk.hr/bugari/de-itineribus-in-turciam-libellus/

Encyclopaedia Britannica. "Standard languages." Accessed January 14, 2020. https://www.britannica.com/topic/dialect/Standard-languages#ref393732

Filipović, Rudolf. 1986. "The Beginnings of Lexicography in Croatia." In *The History of Lexicography,* edited by Reinhard Hartmann, 65-74. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Fontenelle, Thierry. 2015. "Bilingual Dictionaries: History and Development; Current Issues." In *The Oxford Handbook of Lexicography*, edited by Philip Durkin, 44-61. Oxford: Oxford University Press.

Franolić, Branko. 1976. "Was Faust Vrančić the first Croatian lexicographer?" *Annali Istituto Universitario Orientale* 19: 177-182.

Franolić, Branko. 1983. "An historical outline of Croatian lexicography." *Die Welt der Slaven* 7, no. 2: 286-306.

Gabrić-Bagarić, Darija, Marijana Horvat, Ivana Lovrić Jović, and Sanja Perić Gavrančić. 2011. *Jakov Mikalja – Blago jezika slovinskoga (transkripcija i leksikografska interpretacija)*. Zagreb: Institut za hrvatski jezik i jezikoslovlje.

Gadžijeva, Sofija, Ana Kovačević, Milan Mihaljević, Sandra Požar, Johannes Reinhart, Marinka Šimić, and Jasna Vince 2014. *Hrvatski crkvenoslavenski jezik*. Zagreb: Staroslavenski institut – Hrvatska sveučilišna naklada.

Galabov, Ivan Petkov. 1973. *Das Altbulgarische und das Latein im europäischen Mittelalter: zur Problematik der übernationalen Kultursprachen; Antrittsvorlesung gehalten am 14. Dez. 1971 an d. Univ. Salzburg. Salzburg.* Salzburg: Verlag Anton Pustet.

Gelenius, Sigmund. 1537. *Lexicon symphonum quo quatuor linguarum Europae… concordia… indicatur.* Basel: Hieronymus Froben.

Gortan, Veljko, and Vladimir Vratović. 1969-1970. *Hrvatski latinisti.* Zagreb: Matica hrvatska.

Gostl, Igor. 1955. "*Liber de simplicibus*, prvi hrvatski rukopisni, višejezični, terminološki i slikovni rječnik?" *Filologija,* 24-25: 151-156.

Hanks, Patrick. 2013. “Lexicography from Earliest Times to the Present.” In *The Oxford Handbook of the History of Linguistics*, edited by Keith Allan, 503-536. Oxford: Oxford University Press.

History of Dutch Language. “What is a standard language?” Accessed January 14, 2020. http://www.ucl.ac.uk/dutchstudies/an/SP_LINKS_UCL_POPUP/SPs_english/language_history/pages/stand_lang.html

Horvat, Vladimir. 1999. *Bartol Kašić, otac hrvatskoga jezikoslovlja.* Zagreb: Hrvatski studiji.

Jonke, Ljudevit. 1971. “‘Dikcionar’ Fausta Vrančića.” In *Faust Vrančić: Dictionarium quinque nobilissimarum Europae linguarum – Latinae, Italicae, Germanicae, Dalmatiae, & Ungaricae*. Zagreb: Novi Liber.

Jovanović, Neven, Johanna Luggin, Luka Špoljarić, and Lav Šubarić (eds.). 2018. *Neo-Latin Contexts in Croatia and Tyrol: Challenges, Prospects, Case Studies.* Vienna: Böhlau Verlag.

Jovanović, Neven. 2017. “Croatian Neo-Latin Literature and its Uses.” In *A Handbook to Classical reception in Eastern and Central Europe*, edited by Zara Torlone Martirosova, Dana LaCourse Munteanu and Dorota Dutsch, 35-45. Oxford: Wiley–Blackwell.

Kašić, Bartol. 1990. *Hrvatsko-talijanski rječnik s konverzacijskim priručnikom*. Zagreb: Kršćanska sadašnjost – Zavod za jezik Instituta za filologiju i folkloristiku.

Katičić, Radoslav. 2005. “Lodereckerov rječnik – leksikografsko djelo i kulturni spomenik.” *Petar Loderecker: Dictionarium septem diversarum linguarum.* 17-49. Zagreb: Novi Liber.

Keipert, Helmut. 1987. “Kirchenslavisch und Latein. Über die Vergleichbarkeit zweier mittelalterlicher Kultursprachen.” In *Sprache und Literatur Altrußlands: Aufsatzsammlung,* edited by Gerhard Birkfellner, 81-109. Münster: Aschendorff.

Keipert, Helmut. 1996. “Das Lateinische in der Geschichte der russischen Sprache.” In *Eurolatein*, edited by Horst Haider Munske, Alan Kirkness, 106-128. Tübingen: Max Niemeyer Verlag.

Letts, Malcom. (ed.) 2010. *The Pilgrimage of Arnold von Harff, Knight, from Cologne Through Italy, Syria, Egypt, Arabia, Ethiopia, Nubia, Palestine, Turkey, France and Spain, which he Accomplished in the years 1496-1499*. London: Hakluyt Society. Accessed January 14, 2020. https://archive.org/details/in.ernet.dli.2015.59276/page/n5/mode/2up

Loderecker, Petar. 2005. *Dictionarium septem diversarum linguarum.* Zagreb: Novi Liber.

Marti, Roland W. 1987. “Archaismen im Altrussischen.” In *Sprache und Literatur Altrußlands: Aufsatzsammlung,* edited by Gerhard Birkfellner, 111-124. Münster: Aschendorff.

Megiser, Hieronymus. 1592. “Dictionarium quatuor linguarum, videlicet Germanicae, Latinae, Illyricae (quae vulgo Sclavonica appelatur) et Italicae seu Hetruscae.” Graz: Johannes Faber. Accessed January 14, 2020. https://books.google.hr/books?id=cTw8AAAAcAAJ&printsec=frontcover&hl=hr#v=onepage&q&f=false

Mihaljević, Ana. 2020. *Hrvatskoglagoljični tekstovi prevedeni s latinskoga*. Zagreb: Hrvatska sveučilišna naklada – Staroslavenski institut.

Mihaljević, Milan, and Johannes Reinhart. 2005. “The Croatian Redaction: Language and Literature.” *Incontri linguistici* 28: 31-83.

Mihaljević, Milan. 2009. “Hrvatski crkvenoslavenski jezik.” In: *Povijest hrvatskoga jezika, 1. knjiga: srednji vijek*, edited by Ante Bičanić, 283–351. Zagreb: Croatica.

Mihaljević, Milan. 2011. “Položaj crkvenoslavenskoga jezika u hrvatskoj srednjovjekovnoj kulturi.” In Mikalja, Jakov. *Zbornik na trudovi od Megjunarodniot naučen sobir Sveti Naum Ohridski i slovenskata duhovna, kulturna i pismena tradicija (organiziran po povod 1100-godišninata od smrtta na sv. Naum Ohridski)*, edited by Ilija Velev, Aco Girevski, Liljana Makarijoska, Ilija Piperkoski, Kostadina Mokrova, 229–238. Skopje: Univerzitetot Sv. Kiril i Metodij.

Mikalja, Jakov. 2011. *Blago jezika slovinskoga*. Zagreb: Institut za hrvatski jezik i jezikoslovlje.

Musulin, Stjepan. 1959. “Hrvatska i srpska leksikografija.” *Filologija* 2: 41-63.

PHRB: “Portal hrvatske rječničke baštine – Croatian Old Dictionary Portal.” Accessed January 14, 2020. http://crodip.ffzg.hr/

Putanec, Valentin. 1979. “Talijansko-hrvatski i hrvatsko-talijanski rječnik Petra Lupisa Valentiana (Ancona 1527).” *Filologija* 9: 101-138.

Putanec, Valentin. 1982. “Aneksni rječnici u djelu ‘Pokorni i mnogi ini psalmi Davidovi’ (1582) Šimuna Budinića i ‘Pištole i evangelja’ (1613-1857) Ivana Bandulavića.” In *Leksikologija i leksikografija,* edited by Darinka Gortan-Premk, and Drago Ćupić, 227-247. Beograd – Novi Sad: Institut za južnoslavenske jezike Filozofskog fakulteta u Novom Sadu.

Samardžija, Marko. 2005. “Hrvatski jezik u rječnicima F. Vrančića i P. Lodereckera.” In *Petar Loderecker: Dictionarium septem diversarum linguarum.* 51-75. Zagreb: Novi Liber.

Samardžija, Marko. 2011. “Hrvatski leksik i leksikografija.” In *Povijest hrvatskoga jezika 2. knjiga – 16. stoljeće,* edited by Ante Bičanić, Radoslav Katičić, and Josip Lisac, 441-460. Zagreb: Croatica.

Schaffenrath, Florian, and Alexander Winkler (eds.). 2018. *Neo-Latin and the Vernaculars: Bilingual Interactions in the Early Modern Period.* Leiden: Brill.

Snell-Hornby, Mary. 1986. “The Bilingual Dictionary—Victim of its own Tradition?” In *The History of Lexicography,* edited by Reinhard R. K. Hartmann, 207-218. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Stankiewicz, Edward. 1984. *Grammars and Dictionaries of the Slavic Languages from the Middle Ages up to 1850.* Berlin: Walter de Gruyter & Co.

Vajs, Nada. 1995. “Prinos hrvatskoj fitonimiji: Nicolo Roccabonella (1386-1459), Liber de simplicibus.” *Folia onomastica Croatica,* 4: 175-187.

Verbeke, Demmy. 2015. “Neo-Latin’s Interplay with Other Languages.” In: *The Oxford Handbook of Neo-Latin,* edited by Sarah Knight, and Stefan Tilg. Oxford: Oxford University Press.

Vončina, Josip. 1979. “Vrančićev rječnik.” *Filologija,* 9: 7-36.

Vratović, Vladimir. 1989. *Hrvatski latinizam i rimska književnost.* Zagreb: Nakladni zavod Matice hrvatske.

Vratović, Vladimir. 1996. *Hrvati i latinska Europa.* Zagreb: Matica hrvatska.

Vrančić. Faust. 1971. *Dictionarium quinque nobilissimarum Europae linguarum – Latinae, Italicae, Germanicae, Dalmatiae, & Ungaricae.* Zagreb: Novi Liber.

Chapter Fourteen

A Select History of Modern Russian-Tajik Special Lexicography

Abdusalom Mamadnazarov

Abstract

In Tajik bilingual lexicography, more than eighty Russian-Tajik special dictionaries have been created. The aim of this paper is to study and analyze the history, macro- and microstructure of modern Russian-Tajik special lexicographical works published over the eighty-year period since the appearance of the first *Russian-Tajik Dictionary of Biological Terms* in 1941 to the present time. The history of compiling modern bilingual Russian-Tajik dictionaries can be divided into two periods: the Soviet period (1941-1990) and the period of Independence (1991-2019) after the collapse of the Soviet Union and Tajikistan becoming an independent nation.

From the macrostructural point of view, Russian-Tajik dictionaries of the earlier period have the following macrostructural inadequacies:

1. They do not have an Introduction;
2. They do not have a list of references;
3. They do not refer to the earlier published special dictionaries in the field;
4. The author or the authors consider their publication as the first attempt to compile this or that reference book.

From the microstructural point of view, we divide Russian-Tajik bilingual terminological dictionaries into three main types:

1. Translation dictionaries—dictionaries in which Russian terminology and their Tajik equivalents or translations are given

without interpretation of the concept of the term. The majority of Russian-Tajik dictionaries or about 75% referred to are made using this method.

2. The dictionaries in which Russian terminologies are given without interpretation, and their Tajik equivalents and translations are given with interpretation. The number of such dictionaries is about 20%.
3. Dictionaries in which both Russian terms and their Tajik equivalents and translations are given with interpretation. The number of such dictionaries is about 5%.

One of the most important and difficult aspects of compiling bilingual Russian-Tajik terminological dictionaries is the choice of entry words. Analysis of existing dictionaries shows that the choice of the entry words in most cases does not meet the modern requirements of lexicography, resulting in the majority of dictionaries having some failings in content and structure which are underlined in this article.

1. Introduction

During the seventy years of the existence of the Soviet Union, the Russian language served as *lingua franca* for the national republics, and a great number of general active and passive bilingual dictionaries were created for the languages of the nations of the country. Since Russian was the language of instruction at educational institutions, for media presentations—press, TV and radio—and government agencies in general and in most cases, it served as the language of communication for many interrelations among the populations of the republics. In addition, special bilingual dictionaries were compiled to translate special terminology into the languages of these republics. As a result of this government language policy, the nations comprising the Soviet Union became bilingual—speaking Russian, as the state language, and their native, in our case Tajik, language.

With the collapse of the Soviet Union, the situation changed rapidly, and the influence of the Russian language decreased substantially. In the case of Tajikistan, the civil war (1991-1996) forced most of the Russian-speaking population to leave the country for Russia. On the other hand, the Tajik government reduced the number of classes for the study of Russian language at schools and universities, and Russian literature as a subject was removed from the school curriculum. However, the existing socio-economic and cultural relations with Russia still requires a good knowledge of Russian, which is gradually being superseded by English. As a result of the new political and social developments, more attention is now being given to studying English and to English-Tajik lexicography, including English-Tajik bilingual terminological dictionaries, which is out of the scope of study of our article.

With the development of science and technology, new words and expressions are included in the language not as fast as we would like, and are not always reflected in the monolingual and bilingual dictionaries. Therefore, for the standardization, interpretation and translation of special terminology in a particular area, special terminological monolingual, bilingual and less frequently multilingual dictionaries are created. In the age of rapid development of scientific and technological progress, the creation of special Russian-Tajik dictionaries was given much attention in the first period, in Tajik lexicography, since it is through the Russian language, as well as other developed languages, such as English, German, French and others, that a huge flow of information so necessary for specialists enters the Tajik language through Russian. These include special scientific publications, materials of international conferences, patents, Internet resources and other sources. That is, a huge number of special words and expressions and terminology have penetrated into the Tajik language by borrowing from other languages, thus enriching the vocabulary of the Tajik language. However, every action has a certain limit. According to the majority of Tajik linguists, in the first years of the Soviet period the words and expressions of the Russian language were borrowed, very often, without any effort to find their translation or equivalents.

As a result, such words as: *arkhiv (*archives), *sector (*sector), *institute (*institute), *material (*material), *termin (*term), *revolutsiya (*revolution), *economica (*economics), *original (*original), *prosent (*percent) and many, many others were included in Tajik monolingual and bilingual dictionaries, at the same time existing in Tajik equivalents of these words—*boygoni (*archives), *shu'ba (*sector), *donishkada (*Institute), *mavod (*material), *istiloh (*term), *inqilob (*revolution), *iqtisodiyot (*independence), *asli (*original), *foiz (*percent)—were ignored or rarely used.

In special dictionaries published over the past eighty years, there have been attempts to correct such shortcomings and to revive the use of native Tajik words and expressions. On the other hand, at the present stage, many authors of books and articles, especially in the periodical press, fall into another extreme, discarding the already known Russian borrowings, and use a large number of Arabic words and expressions that are unknown and incomprehensible to the majority of the population, thus, pushing many readers away from understanding and using these publications.

In the history of Tajik bilingual lexicography, special dictionaries can be mainly divided into two types based on the object and purpose of the study: a) special dictionaries in which the Russian language is the object of study, and the Tajik language is the purpose (*more than 80 dictionaries*) and b) special dictionaries in which English is the object of study, and the Tajik language is the purpose (*more than 20 dictionaries*).

Special bilingual or terminological dictionaries, along with general and special explanatory dictionaries, are of great importance for the formation and development of the vocabulary of the Tajik language; they also create the basis for the creation in the future of a large explanatory dictionary of the Tajik language in general, and Russian-Tajik general and special dictionaries, in particular.

The number of specialized bilingual terminological dictionaries in Tajik lexicography has increased substantially, especially in recent decades. Some of them are small publications describing only 200 to 300 words and resemble a simple glossary of bilingual words, and are created mainly for educational purposes; others register a larger number, from five hundred to four to seven thousand or more words and expressions. Most special dictionaries are bilingual, and only a small number are multilingual.

Unfortunately, the critical analysis of the micro- and macrostructure and the content of these dictionaries in Tajik lexicography is given insufficient attention by lexicographers. In this article, we will try to study and analyze the micro- and macrostructure of only Russian-Tajik terminological lexicographic works published over the past eighty years and would like to bring some clarity to this issue.

In Tajik lexicography, four words are used as the synonym of the word "*dictionary*"—*farhang*, *lughat*, *qomus*, and in recent years the word *loghtanoma* is also used. We prefer and recommend the term *lughat* as the most appropriate. As to the word *farhang*, this word has a broader meaning and can be applied to large monolingual explanatory and general bilingual dictionaries.

Of course, within the frame of a single article it is not possible to consider and analyze all available dictionaries; however, we will try to analyze the most important bilingual Russian-Tajik terminological dictionaries.

2. The Soviet Period

The creation of Russian-Tajik terminological dictionaries began in the 1940s for good reason, as the main source of information in the Soviet period was the Russian language, and, since then, dozens of Russian-Tajik terminological dictionaries were created which cover the terminology of various branches of science and technology, ranging from terminology in chemistry, biology, economics and law to the terminology of art, agriculture, etc.

In the 1940s, the first steps to create special dictionaries, which is the initial period in the formation of bilingual Russian-Tajik special lexicography, were made. Among these we can mention: the *Russian-Tajik Dictionary of Biology* by E.A. Fuzaylov and others (1941), *the Russian-Tajik Dictionary* of *Military Terms* edited by E.N. Pavlovsky (1942), *the Russian-Tajik Dictionary on Chemistry* by M.A. Rajabov and M. Alizoda (1948), *the Russian-Tajik Dictionary of Geographical Names* by H. Rahmatulloev and others (1948).

In the 1960s and 1970s, the process of creating the special dictionaries has been developed actively as the result of which we have: the *Russian-Tajik Dictionary on Mathematics* by H. Muhammedova (1960), the *Russian-Tajik Dictionary of Physical Terms* by W.H. Sodiqov (1960), the *Russian-Tajik Dictionary of Philosophical Terminology* by M. Asimov, M. Bachev and M. Dinorshoev (1966), the *Russian-Tajik Dictionary of Chemical Terminology* by L. Rajabov and A.R. Rakhmonov (1967) and others.

In the 1980s, one can notice a growing interest in the creation of special dictionaries for construction, and architecture, as well as scientific and technical dictionaries, for those industries in which no bilingual dictionaries had been created before, among them: the *Russian-Tajik Dictionary of Construction Terms* by N. Sharopov, A. Komilova, T. Juraev (1982), A *Short Russian-Tajik Dictionary of Architectural and Artistic Terms* by R.S. Muqimov (1983) and *Russian-Tajik Dictionary of General Technical Terminology* by T.K. Juraev (1987). Among the most successful bilingual dictionaries compiled on a serious scientific basis one can mention the *Russian-Tajik Dictionary of Physical Terms* by S. Qodiri (1985). During this period, more than sixteen special bilingual dictionaries were published.

3. The Period of Independence

In the 1990s, after Tajikistan became an independent country, the creation of bilingual Russian-Tajik special dictionaries continued.

In 1991, the first *Russian-Tajik Dictionary on Construction and Architecture* was created by R. Ruziev and M.A. Asimov, which was published by the Tajik Technical University. This dictionary describes more than 800 of the most frequently used words and expressions on construction and architecture, borrowed from Russian, English, Italian, Latin, Dutch and other languages through Russian. The compilers do not only translate the terminology of construction and architecture into Tajik and describe them, but also point to the etymology of these words. For example:

> **Avansena**—(from *Fr. avante – pesh and scene* - *sahna*) – *peshsahna, qismi kushodi sahna, ki as parda berun, bayni parteru orkester voqe' ast* (*peshsahna, an open part of the stage, that is not covered by curtains, between partyer and orchestra*) (Ruziev 1991, 4).

The word list of this dictionary is based on textbooks on construction and architecture and is mainly intended for students and teachers of this field who use the literature published in Russian. The compilers do not provide a list of references.

In 1993, the publishing house Simurg published a *Brief Russian-Tajik Dictionary of Office Work* by S. Nazarzoda, which contains four thousand words and expressions used in office work. The purpose of the dictionary is to assist in the correct use of terminology in the field of institutions and enterprises of the country. Russian terminology and their Tajik equivalents or translations in this bilingual dictionary are given without interpretation of the term itself. Since most of the terms of the Tajik language are in the process of formation, the entries in the dictionary contain several Tajik synonyms to the entry word. According to the compiler of the dictionary, providing two or three synonyms as equivalents of one term is aimed at streamlining the terms, so that in the future for the compilation of larger dictionaries different ways of term formation are taken into account (Nazarzoda 1993, 3). The dictionary consists of a brief Preface by the author and a Tajik-Russian glossary. The author does not provide a list of references.

In 1999, Maorif Publisher issued *The Russian-Tajik Dictionary of Commonly Used Words* by G. Juraev, S. Fathulloeva and A. Asadov. The dictionary consists of a brief introduction and the Russian-Tajik dictionary itself. The compilers of the dictionary do not give a list of used literature, but from the introduction it can be concluded that most of the words were borrowed from bilingual, explanatory and terminological dictionaries. That is, the purpose of the dictionary, according to the compilers, is not to create a terminological dictionary of any field, but to create an accessible dictionary (Juraev 1999, 3). We see that the wordlist of a dictionary that was compiled taking into consideration the terminology of agencies and related enterprises. In the introduction, however, how and by what method the authors have identified and selected the most common words are not specified. In the Tajik part of the dictionary, one can notice an intensive use of difficult to understand, even incomprehensible, Arabic words on the one hand, and a selection method that disregards the well-known and most common native Tajik words on the other hand.

The first *Short Russian-Tajik Dictionary on Taxes and Taxation* belongs to S. Saidov's pen, which was published in 2000 by the Sharqi Ozod Publishing House. The compiler of the dictionary, having studied records and management documents in the field of taxes and taxation, collected in his dictionary the most common words of this field. The compiler does not unduly restrict the number of words in the dictionary, however, according to our calculations; it describes more than six thousand words and expressions of the Russian language and their Tajik equivalents. Russian terminology, their Tajik equivalents and translation in this dictionary are given without interpretation of the concept of the term. In compiling the dictionary, special attention was paid to the meaning of words that are used and have a specific meaning in the field of taxes and taxation. There are also words that do not belong separately to this area and only in a combination of words or expressions acquire terminological significance. In these cases, the meanings of such words are given not separately but in phrases and expressions.

This dictionary was compiled with great care. In the introduction, the compiler, continuing the discussions on the theme of borrowing foreign words, correctly notes that, fortunately after the adoption of the law on state language and Tajikistan's independence, many Tajik words were revived again and took their rightful place in the vocabulary of the Tajik language. However, the wave of revival led to another undesirable consequence, as a result of which, instead of Russian-international words, as well as native Tajik words their Arabic unknown and incomprehensible equivalents were used (Saidov 2000, 4). The rules of spelling of the Tajik language, which of course have a certain value, included in the dictionary, however, like many other Tajik lexicographers, the compiler does not list the sources used. It should be emphasized that this dictionary has a certain value in the standardization of terms in this field.

Among the special terminological dictionaries, the dictionary by Y.A. Samarkandi entitled *Russian-Tajik Dictionary of Physical Terms*, published in 2002, occupies a special place in Tajik terminological lexicography. First, this dictionary is the most complete bilingual dictionary in physics, which describes and provides interpretation and translation of more than 33,500 terms, expressions and concepts of physics from Russian into Tajik. Secondly, the compiler of the dictionary gives a large number of expressions using physical terms, as it is through expressions, in most cases, that scientific terms and concepts become more understandable. Alongside new and modern terms, the compiler of the dictionary also cites outdated and out-of-use words-terms, to facilitate the understanding of previously published works on physics. This dictionary is the result of more than 30 years of work by the author. In fact, this dictionary is the third terminological dictionary in Tajik lexicography in the field of physics, after the works by H. U. Sadiqov (1960) and S. Qodiri (1985). The purpose of this work, as well as other terminological dictionaries in this area, is to rationalize the development and use of physical terms in the Tajik language. Although the compiler of the dictionary does not list the literature used, in the introduction he indicates that textbooks on physics of secondary schools and universities, encyclopedias and various scientific journals in physics were used to compile the dictionary.

In 2007, a *Brief Explanatory Russian-Tajik Dictionary of Architecture and Construction* was published by R. Muqimov, T. Juraev and M. Karimov by ICOMOS publications, which describes about three thousand terms used in architecture and construction. This dictionary is the fourth in this field. In terms of macrostructure, the dictionary consists of an Introduction, a list of references (24 titles), the glossary and a large number of drawings, which are included in the Annex. The microstructure of the entry of this guide consists of a word or expression used in architecture and construction in Russian without interpretation, a reference to the language from which the term is borrowed, the Tajik equivalent or translation and interpretation of the terms in Tajik. For example:

> **Arch** (*from Ita. Arco, from Lat. Acrus – toq*) – *ravoq, toq, bolopushi kajkhattai bayni du takyagoh,* mas. *ayvon, pul, va gh.* (*a curved structure that supports the weight of sth above it, such as bridge or upper part of a building*) [Muqimov 2007, 19].

The dictionary contains some entries that are not terms of construction and architecture, but have descriptions of an encyclopedic character, such as: *Adzhinatepa*, Architecture of Tajikistan, Bactria, Dushanbe, etc. Although the authors provide a fairly large list of used literature from the field of architecture and construction, previous dictionaries of this area 9for example, Ruziev’s dictionary, etc.) are not evaluated and are not considered critically, and are not included in the list of used literature.

The Russian-Tajik Economic Dictionary by S. Habibov, H. Faqerov and M. Jamshedov was first published in 1999 and then with changes and additions in 2010 by Irfon Publishing House. The dictionary is quite complete and covers more than fifteen thousand of the most important words and expressions in the field of economics and mainly describes and gives a brief interpretation of the terminology of the market economy. It should be noted that this dictionary is the most complete Russian-Tajik economic dictionary ever published in the Republic of Tajikistan and can serve as an important and solid basis for the creation of a full Russian-Tajik economic dictionary in the future. From a macrostructural point of view, the dictionary consists of an Introduction by the compilers and the Russian-Tajik word list. The list of the used literature is given in the introduction and consists of only four items. The compilers of the dictionary did not use previously published explanatory Tajik and bilingual economic dictionaries (1973, 1979, 1998 and 1999). Russian headwords are given without interpretation, and their Tajik equivalents are translated and given with interpretation.

The first *Russian-Tajik Dictionary of Animal Husbandry* (400 p.) in the Tajik bilingual lexicography, published in 2011, belongs to T.A. Qadirov and K.A. Dustov. This dictionary gives translation of the entries and interpretation in the Tajik language for more than 800 terms related to livestock. Russian terminology is given without interpretation in this explanatory dictionary, and their Tajik equivalents and translations are interpreted in detail. The interpretation of most terms takes almost half a page and carries information of an encyclopedic nature. At the end of the dictionary, an index of terms is given with the appropriate page number. The list of references consists of 25 reference works.

In our review of terminological dictionaries, we included a *Brief Dictionary of Terms and Concepts of Journalism* (114 p.), compiled by M. Murodov and published in 2011. If we look at the name of the dictionary, it may seem that the dictionary is a monolingual terminological dictionary in the field of journalism, but in fact, it is a bilingual Russian-Tajik dictionary that describes more than one thousand terms and concepts of journalism, translated from Russian into Tajik. That is, this dictionary can be attributed to the type of Russian-Tajik dictionaries where the terms of the Russian language are given without interpretation, and their Tajik equivalents and translation with interpretation and explanation. The dictionary is based on the Russian alphabet and it would be more accurate to call this dictionary a *Brief Russian-Tajik Dictionary of Terms and Concepts of Journalism*. In terms of its microstructure, a dictionary entry consists of a Russian term, a reference to the etymology of the words in brackets (from English, Latin, German, etc.), Tajik equivalent or translation and interpretation of the terms in Tajik.

In 2011, the first *Russian Tajik Dictionary on Metrology, Standardization and Certification of Goods and Services* (160 p.) compiled by I.M. Mirzomiddinov and I. Ganiev was published which describes more than 1,300 words and expressions used in this field. It is known that most of the terms in this field are borrowed from well-known international languages and their use in the Tajik language is associated with certain difficulties. According to the compilers of the dictionary, this happens because many authors consider these words borrowed from Russian, and try to replace them with Persian and Arabic words. As a result, in the Tajik language there appear words that are little known and do not express the terminological meaning of the original words (Mirzomiddinov 2011, 5). Therefore, the main purpose for compiling this dictionary, according to the authors, is the interpretation of international words, and their translation and dissemination in the Tajik language. The dictionary consists of an Introduction, a section on the structure of the dictionary, the dictionary itself, an Appendix (Mirzomiddinov 2011, 116-158) and a list of references (12 titles). The dictionary entry consists of a Russian headword with a capital letter, an indication on the etymology of the word, an interpretation of the main and figurative meanings and an expression with the specified term. For example:

> **Magnet** (*Gr. magnetis from Magnetis lithos, a stone from Magnesia*, Shahri Kadima dar Osiyi Khurd (*an ancient city in Minor Asia*) 1. *physics*. *ohanrabo, magnit* (jisme, key maydoni magniti ba vujud movarad – *a piece of iron that creates magnetic field*) [Mirzomiddinov 2011, 80].

The first *Russian-Tajik Dictionary on Agricultural Machinery and Technology of Metals* was published in 2005 and with additions and changes (265 p.) in 2011 by H. Safarov; it describes more than 7,500 words and expressions and is the most complete terminological dictionary in the bilingual Tajik technical lexicography. The dictionary is primarily recommended for students of higher educational institutions, especially for agricultural and engineering faculties. In terms of its macrostructure, the dictionary consists of the publisher's Introduction, the author's Preface, the dictionary itself and the list of references [16 items] and is a type of dictionary where Russian terms are given without interpretation, and their Tajik equivalents or translations with interpretation. The compilers created this dictionary taking into account the training programs on the use of machine-tractor parks and metal technology, textbooks and especially technical specifications, state standards and other technical periodicals.

The first *Russian-Tajik Dictionary of Ichthyology* (588 p.), compiled by Rasulov, was published in 2005. The dictionary is devoted to one of the main sections of biology—*ichthyology*, which is of great practical importance for various sectors of the economy and contributes to the scientifically based development of fisheries. The dictionary is a bilingual explanatory dictionary and a kind of reference book of encyclopedic nature, where some articles are provided with photographs and diagrams, and consists of an Introduction in Tajik and Russian, explanatory dictionary both in Russian and Tajik languages and some appendixes (e.g. *summary of water resources of Tajikistan, the composition and types of fish and water bodies of the country*). The list of sources (38 items) is given in the introduction with indication only of the surnames of authors and year of the edition without indications of names of sources. For example: (Alexandrov, Pozhidaev, 2006; Aliyev, etc. 1988; Veselov, 1977), etc. It would, of course, be desirable and necessary to specify the names of directories.

The Tajik-Russian special lexicography was not part of the scope of this article, but since T. Shokirov's *Tajik-Russian Dictionary of Legal Terms* (p. 368) is the only reference work (as far as we know), we decided to say a few words about it. This *Tajik-Russian Dictionary of Legal Terminology* is the second Tajik-Russian terminological dictionary in the history of this field (the first one was compiled by S.A. Rajabov, A. L. Bukharizade et al., in 1965). This dictionary is one of the most complete bilingual terminological dictionaries, in which more than nine thousand words and terminology of various fields of law are translated from Tajik into Russian and, according to the author, is intended for students of lawyers, political scientists, civil servants, law enforcement and military personnel. Unfortunately, the dictionary is printed in only 100 copies, which does not justify the purpose defined by the compiler. In our opinion, this dictionary is most likely to benefit translators who are engaged in translation from Tajik into Russian, the number of which is small.

Practically, this dictionary is the fifth special dictionary in this field and the fourth bilingual terminological dictionary in the history of Tajik language after *Tajik-Russian Dictionary* by S. Rajabov and Bukharizade (1965), *Russian-Tajik Dictionaries* by R. Abdulloev (1974) and by Sh. Rosiqov, F. Tokhirov, T. Sharipov (1994). It should be noted that the author mentions only Abdulloev's dictionary. The macrostructure of the dictionary consists of an introduction, the dictionary itself and a list of references [31 titles], where Tajik and Russian terms are given without interpretation of the term. Of course, the dictionary can contribute to the standardization and streamlining of vocabulary in this area.

4. Conclusions

One of the most important and difficult aspects of compiling bilingual Russian-Tajik terminological dictionaries is the choice of headwords. Analysis of existing dictionaries shows that the choice of the entry in most cases does not meet the modern requirements of lexicography, with the result that the majority of the dictionaries have some shortcomings in content and structure. Of these, the most significant are:

1. The dictionaries, along with the terminology of particular area, also include words and expressions of common sense that are not related to the terms of the field.
2. Terminological dictionaries do not always reflect all the special terms of a particular field.
3. The special terminology of a particular field in the dictionaries of different authors is translated into Tajik in different ways, or their interpretation differs from each other.

It means that the process of unification and standardization of terminologies is still underway in Tajik language.

Some of the dictionaries of the earlier period have the following macrostructural weaknesses:

1. They do not have an Introduction;
2. They do not have a list of references;
3. They do not refer to the earlier published special dictionaries in the field;
4. The author or the authors consider their publication as the first attempt to compile this or that reference book. The reason for no. 4, we believe, may have been:
    a) The author/s did not have excess to the earlier published dictionaries (not all of the publications can be found in the libraries); or
    b) The author/s did not make necessity efforts to find relevant information, dictionaries; or
    c) The author/s purposefully mislead the readers in order to emphasize that their publication is the first in the field, perhaps out of pride. And/or other reasons that we may not be aware of.

From a microstructural point of view, we divide Russian-Tajik bilingual terminological dictionaries into three main types:

1. Translation dictionaries—dictionaries in which Russian terminology and their Tajik equivalents or translations are given without interpretation of the concept of the term. The majority of

Russian-Tajik dictionaries or about 75% referred to are made using this method.

2. The dictionaries in which Russian terminologies are given without interpretation, and their Tajik equivalents and translations are given with interpretation. The number of such dictionaries is about 20%.
3. Dictionaries in which both Russian terms and their Tajik equivalents and translations are given with interpretation. The number of such dictionaries is about 5%.

The formation and development of Tajik special bilingual dictionaries contribute to standardization of terms in the Tajik language and provide great assistance in the creation of larger bilingual special dictionaries and the corpus of Tajik language in general in the future. Therefore, there is still much work to be done.

Bibliography

Asimov M., Bachaev M., and Dinorshoev M. 1966. *Russian-Tajik Dictionary of Philosophical Terminology*. Dushanbe: Irfon.

Bejoint, Henri. 1994. *Tradition and Innovation in Modern English Dictionaries*. Clarendon Press. Oxford.

Bejoint, H. 2000. *Modern Lexicography: An Introduction*. Oxford University Press.

Fuzaylov, E.A. 1941. *Russian-Tajik Dictionary of Biology*. Stalinabad: Tadgiz.

Habibov, S., Faqerov, H.N., and Jamshedov, M. 2010. *Dictionary of Economic Terms (Russian-Tajik)*. Dushanbe, IRF.

Hartmann, R.R.K. 1983. “On Theory and Practice.” In *Lexicography: Principle and Practice*, edited by R.R.K Hartmann and K.A. Hartmann, 3-11. London/New York: Academic Press.

Hartmann, R.R.K. 2001. *Teaching and Researching Lexicography*. Longman.

Juraev, G., Fathulloev, S., and Asadov, A. 1999. *The Russian-Tajik Dictionary of Commonly Used Words*. Dushanbe.

Juraev, T.K. 1987. *Russian-Tajik Dictionary of General Technical Terminology*. Dushanbe: Maorif.

Landau, S.I. 2001. *Dictionaries: The Art and Craft of Lexicography*. 2nd edition. Cambridge University Press.

Mamadnazarov A. 2013. *Comparative Analysis of the Main Dictionaries in Tajik Lexicography*. Dushanbe: Er-graph.

Mamadnazarov A. 2016. *The Formation and Development of Tajik Translation Lexicography in the XX and the Beginning of XXI Centuries*. Monography. Dushanbe: Er-graph.
Mamadnazarov A. 2018. *Oxford English-Tajik Dictionary*. Online at: http://www.tg.oxforddictionaries.com
Mirzomiddinov, I.M., and Ganiev, I. 2011. *Russian-Tajik Dictionary on Metrology, Standardization and Certification of Goods and Services.* Dushanbe: Irfon.
Muhammedov, H. 1960. *Russian-Tajik Dictionary on Mathematics*. Stalinabad: Tadgiz.
Mukimov, R.S. 1983. *A Short Russian-Tajik Dictionary of Architectural and Artistic Terms*. Dushanbe: TTU.
Muqimov, R., Juraev, T., and Karimov M. 2007. *Brief Russian-Tajik Dictionary of Architectural and Artistic Terms.* Dushanbe: ICOMOS.
Murodov, M. 2011. *A Brief Dictionary of Terms and Concepts of Journalism*. Dushanbe: Shujoiyon.
Nazarzoda, S. 1993. *A Brief Dictionary of Office Terms (Russian-Tajik).* Dushanbe: Irfon.
Pavlovsky, E.N. 1942. *Russian-Tajik Dictionary of Military Terms*. Stalinabad: Tadgiz.
Qadirov, T.A., and Dustov, K.A. 2011. *Russian-Tajik Dictionary of Animal Husbandry*. Dushanbe: Adib.
Qodiri, S. 1985. *Russian-Tajik Dictionary of Physical Terms*. Dushanbe: Maorif.
Rahmatulloev, H. 1948. *Russian-Tajik Dictionary of Geographical Names*. Stalinabad: Tadgiz.
Rajabov, L., Alizoda M. 1948. *Russian-Tajik Dictionary on Chemistry*. Stalinabad.
Rajabov, L., Rakhmonov A. R. 1967. *Russian-Tajik Dictionary of Chemical Terminology*. Dushanbe: Irfon.
Rasulov A., 2011. *Russian-Tajik Dictionary of Ichthyology*. Dushanbe.
Ruziev, Ch. R., and Asimova, M.A. 1991. *Dictionary of Construction and Architecture* (*Russian-Tajik).* Dushanbe: DTT.
Safarov, H. 2011. *Russian-Tajik Dictionary of Agricultural Machinery and Technology of Metals (Russian-Tajik).* Dushanbe: Adib.
Saidov, S. 2000. *A Short Russian-Tajik dictionary on Taxes and Taxation*. Dushanbe: Sharqi Ozod.
Samarkandi, J. S. 2002. *Russian-Tajik Dictionary of Physical Terms*. Dushanbe: Amri ilm.
Sharopov, N., Komilova, A., and Juraev, T. 1982. *Russian-Tajik Dictionary of Construction Terms*. Dushanbe: Irfon.

Shokirov, T.S. 2011. *Tajik-Russian Dictionary of Legal Terminology*. Khujand: Nargis.

Sodiqov, H.U. 1960. *Russian-Tajik Dictionary of Physical Terms*. Stalinabad: Tadgiz.

CHAPTER FIFTEEN

LEXICOGRAPHY OF FRISIAN

FRITS VAN DER KUIP, JANNEKE SPOELSTRA AND WILLEM VISSER

Abstract

In this article, we give a brief outline of the lexicography of Frisian, one of the West Germanic languages. Frisian is spoken in the Dutch Province of Fryslân (Friesland). Its minority language status is reflected in its lexicographical output. As with many other minority languages, this output mainly consists of practical, bilingual dictionaries. Dutch having been the dominant language in Fryslân for a long time, these are of course mostly Frisian-Dutch and Dutch-Frisian dictionaries. While a scholarly dictionary of Frisian may seem less obvious, it does in fact exist: a 25-volume, scholarly dictionary describing the Frisian language between 1800 and 1975. Plans have been made to compile scholarly dictionaries for Frisian in other periods, but so far these have not materialized. We are currently working on the Online Dutch-Frisian Dictionary.

1. Introduction

Frisian is one of the smaller branches of the Germanic language tree and belongs to the West Germanic language group, along with English, German and Dutch. In the Middle Ages, Frisian was spoken in the northwest of the European continent, along the North Sea coast from the Netherlands to Denmark. Nowadays, Frisian is divided into three main dialects: West Frisian, East Frisian and North Frisian. These languages share several grammatical and lexical characteristics, but are not mutually intelligible, or only to a very limited extent.

West Frisian (hereafter called 'Frisian') is the mother tongue of 57.3% of the population of the Dutch Province of Fryslân, which has a total population of approximately 650,000 (Fryske Taalatlas 2020). After ratification of the European Charter for Minority Languages in 1996, Frisian gained legal recognition as the Netherlands' second national language. Such recognition means that the authorities have an obligation to maintain the language. The Use of Frisian Act in 2014 gave Frisian citizens the right to use their own language, whether Dutch or Frisian, in court and in their communication with administrative bodies within the Province of Fryslân. East and North Frisian are spoken in a few areas of northwest Germany, albeit by a small minority. The various North Frisian dialects have nearly 10,000 speakers in total, East Frisian has less than 1,000.

Like many other languages, Frisian can be divided into different language periods. Traditionally, the following periodization was used, based on historical, literary grounds:

- Old Frisian (before 1550)
- Middle Frisian (1550-1800)
- Modern Frisian (from 1800 onwards)

Today, linguists use a periodization based on linguistic grounds—which we need not explain further here—resulting in the following typological division (and further subdivision):

- Old Frisian (before 1450)
- Middle Frisian (1450-1550)
- Modern Frisian (from 1550 onwards)
  - Early Modern Frisian (1550-1800)
  - Modern Frisian (from 1800 onwards)

Where we refer to period dictionaries in this article, we mean the *traditional* periodization.

2. Lexicography of Frisian

The Fryske Akademy (Frisian Academy) was founded in 1938. It has its seat in Leeuwarden, the capital of Fryslân. One of the main objectives of the Fryske Akademy was, and still is, the codification and elaboration (as defined by Einar Haugen, 1966a,b) of the Frisian language. Not coincidentally, the Akademy's first scholarly staff member was a lexicographer and the Akademy has pursued a two-track policy from the beginning: scholarly lexicography on the one hand, and practical lexicography on the other. The Fryske Akademy must meet not only the demands of scholarship but also the needs of the Frisian (language) community from which it originates, in which it is rooted and which it 'serves'. Financial support by the provincial government, which pursues an active language policy, is essential when compiling concise dictionaries for a relatively small number of speakers of a minority language such as Frisian, as such publications are scarcely profitable and therefore not interesting to commercial publishers.

2.1 Contemporary lexicography

2.1.1 Lexicon Frisicum (1872)

Joost Halbertsma (1789-1869), a vicar and a scholar, compiled the first scholarly dictionary of Frisian, including all historical, contemporary and regional varieties: Old Frisian, Middle Frisian and Modern Frisian, as well as North Frisian, East Frisian and their dialectal varieties. He wrote for an international scholarly audience and used Latin as the dictionary's metalanguage. When he died, the dictionary's manuscript had reached the entry *Feer*. Halbertsma's dictionary is not strictly alphabetical, however, which is why it also contains articles that come after *Feer* alphabetically. Three years after his death, Halbertsma's son published the completed section as *Lexicon Frisicum. A-Feer* (Halbertsma 1872; Dykstra 2011).

2.1.2 Friesch Woordenboek (1900-1911)

In 1879, the Provincial Executive of Fryslân decided there should be a comprehensive dictionary of Modern Frisian, taking as its point of departure the *Lexicon Frisicum* and the dictionary material that Halbertsma had left to the Province of Fryslân. This new dictionary, the *Friesch woordenboek*, focused primarily on language practice and had Dutch as its metalanguage. It was compiled by Waling Dykstra (1821-1914)—not a linguist but a writer, journalist and playwright with a keen eye for Frisian idiom—with the assistance of others. The dictionary appeared in three volumes (1900, 1903 and 1911). Its subtitle refers to the *Lexicon Frisicum* on which it was based. The three volumes were complemented by the *Friesche Naamlijst* (*Onomasticon Frisicum*) (Winkler 1898).

2.1.3 Nederlandsch-Friesch Woordenboek (1918)

Early in the twentieth century, a growing number of Frisians wished to write in their mother tongue. Frisian was not yet included in the school curriculum and there was a need for a Dutch-Frisian production/translation dictionary. The librarians—again not linguists—G.A. Wumkes and A. de Vries compiled the first Dutch-Frisian dictionary (Wumkes and De Vries 1918), working from the national language to the minority language. Frisian was relatively early in this regard, compared with Low German and Low Saxon, for example (Popkema 2010, 83).

Not everyone was pleased with this initiative. Even before the dictionary was published, the dialectologist Jan Jelles Hof (1916, 317) commented that the Frisian users of the dictionary apparently needed it because they did not have enough knowledge of Frisian and were used to thinking in Dutch when writing in Frisian:

> *'It Nederlânsk hat ús sa yn 'e macht, en wy kenne as 't knypt sa'n bytsje Frysk, det wy moatte allegedurigen neisjên kinne hwet it Fryske wird is for 't Nederlânsk dat him yn ús brein foarmet as wy tinke. Oars kinne wy net Frysk skriuwe.'*
> ['Dutch has us so much in its sway and we know so little Frisian when push comes to shove that we constantly need to check what the Frisian word is for the Dutch that forms in our brain while we are thinking. Otherwise we cannot write Frisian.']

This state of affairs, which Hof mentioned with irony at the time, now appears to have increasingly become reality. In 1976, the Frisian linguist Sjölin noted with regard to Frisian-Dutch bilingualism that there was an extraordinary degree of similarity between the two languages' phonetics, phonology, morphology and syntax. It was very likely possible, he thought, to describe both languages with the same grammar, with Frisian splitting off from Dutch only at the morphophonemic level by virtue of a limited number of conversion rules (Sjölin 1976, 11). Sjölin's hypothesis appears to be confirmed by recent research by Bosma and Nota suggesting that bilingual children—with Frisian and Dutch as their languages—use their dominant reading language, Dutch, to read their non-dominant reading language, Frisian, but conversely, do not use Frisian to read Dutch (Bosma and Nota 2019).

2.1.4 Wurdboek fan de Fryske taal (1984-2011)

The most important lexicographic task of the Fryske Akademy was to set up, compile and publish the *Wurdboek fan de Fryske taal/Woordenboek der Friese taal* (Dictionary of the Frisian language, hereafter 'WFT'). It is a scholarly lexicographic description of Modern Frisian between 1800 and 1975, following the example of the *Woordenboek der Nederlandsche Taal* (Dictionary of the Dutch language, hereafter 'WNT'). The start of the massive WFT project coincided with the founding of the Fryske Akademy in 1938. A bare bones scholarly staff, assisted by many volunteers, began by establishing an index system of cards bearing quotations from written and oral sources. The written sources included earlier dictionaries.

Unfortunately, the lexicographers discovered that the card index was biased towards 'interesting' words and idioms. They therefore felt the need to establish a second card index system with a better representation of ordinary words, especially function words. The second card index system, created with stencils and scissors and again with the help of an army of volunteers, was very modern for its time and can be regarded as a language database *avant la lettre*. It has, for example, a couple of drawers of quotations using the definite article *de* (the), the most frequent word in Frisian, whereas the first card index system has only a few index cards on *de*.

Because Dutch has been dominant in many domains, Frisian has a large number of lexical gaps. To close these gaps, the Fryske Akademy's lexicographers enlisted the help of a number of specialists in various ancient and modern disciplines. In time, it also increased its staff to six lexicographers and a number of assistants and proceeded to publish a volume of the dictionary every year from 1984 to 2011.

As Frisian is a minority language, the WFT metalanguage is not Frisian but Dutch. Frisian specialists twice debated the choice of metalanguage, putting forward fierce arguments. Ultimately, Dutch was chosen for practical reasons. The WFT differs in this respect from the historical dictionaries of the larger Germanic languages that function as national languages, such as English (*Oxford English Dictionary*, OED), German (*Deutsches Wörterbuch*, DWB) and Dutch (*Woordenboek der Nederlandsche Taal*, WNT).

In retrospect, the decision turned out to be a blessing in disguise. With Dutch as the metalanguage, it was fairly easy to link the WFT to the historical dictionaries of the Dutch language in the Integrated Language Database of the Dutch Language Institute in Leiden (gtb.ivdnt.org).

Queen Beatrix came to Fryslân in 1984 to receive the first copy of the WFT. In 2010, she visited again to launch the online version of the complete WFT during the opening of the fourteenth Euralex Conference. These two occasions illustrate the impact of the computer on the development of lexicography: from writing by hand to data entry, and from analog to digital.

2.1.5 Concise dictionaries

The first concise dictionary, compiled by the editors of the WFT, was the *Frysk Wurdboek* (Frisian Dictionary). The Dutch-Frisian and the Frisian-Dutch sections were published separately at first (Buwalda, Meerburg and Poortinga 1952, 1956) and, from 1971 onwards, in a single bound volume.

In the eighties, the *Frysk Wurdboek* was succeeded by new concise dictionaries. Frisian gained a more important place in education, media and government, requiring more extensive dictionaries. The Frisian-Dutch part of the new *Frysk Wurdboek* was published in 1984 and has 55,000 entries (Zantema 1984). The Dutch-Frisian part was published a year later and has 35,000 entries (Visser 1985). These dictionaries have been reprinted several times.

The most recent Dutch-Frisian dictionary, as mentioned above, was published in 1985. It is therefore outdated and in need of replacement. The *Prisma woordenboek Fries* (Prisma Dictionary of Frisian) (Spoelstra, Post and Hut 2007), published in 2007 and widely used in the meantime, has partly filled this gap, but is too compendious to serve as a full-fledged substitute. The 1985 dictionary has not only become outdated, it is also rather limited in offering examples of usage in specific contexts. An updated, more extensive new dictionary is obviously needed, and more urgently than ever, with Frisian being under increasing pressure from Dutch. The importance of such a dictionary can hardly be overestimated.

2.1.6 Frysk Hânwurdboek (2008), a monolingual intermezzo

The lexicographical emancipation of Frisian reached a climax with the publication of the *Frysk Hânwurdboek* (Concise Frisian Dictionary) with 70,000 entries (Duijff, Van der Kuip, De Haan and Sijens 2008). For the first (and only?) time, the metalanguage was Frisian. The Fryske Akademy made this decision as a political gesture to the ‘Frisian Movement’, an interest group committed to creating more space for the Frisian language and culture. The *Frysk Hânwurdboek* can be seen as a major achievement in Frisian lexicography.

2.1.7 Online Nederlânsk-Frysk Wurdboek

The Fryske Akademy's lexicographers are now working on the *Online Nederlânsk-Frysk Wurdboek* (Online Dutch-Frisian Dictionary, ONFW). The ONFW is primarily a production dictionary with modern standard Dutch as the source language and standard Frisian as the target language. The ONFW's main target group consists of active language users wanting either to translate from Dutch into Frisian or to produce texts in Frisian. The ONFW is not only intended for speakers of Dutch but also, and especially, targets Frisian speakers. Dutch, as the majority language, exerts immense pressure on Frisian, while education in Frisian is far from optimal and certainly not enough to countervail the influence of Dutch. As a consequence, many Frisian speakers mix Dutch words, forms and structures into their language and write Dutch more often, and better, than they do Frisian. In practice, the average Frisian speaker never masters standard written Frisian; one could even say that many Frisians think in Dutch when they write in Frisian, as Hof (1916, 317) noted (see 2.1.3 above), and are illiterate in their own language. Popkema (2010, 89) argues that, for the purposes of text production, Frisian speakers are best served by a dictionary that has Dutch as the source language and Frisian as the target language.

The design of a dictionary depends on its intended users, and in the case of the ONFW, these are active language users wishing to write in Frisian. This group is better served by an illustration of how Frisian words should or should not be used than by a lengthy list of Frisian equivalents of transparent compounds (which does not mean that these do not merit a place in the dictionary). In the ONFW, the emphasis is therefore more on the microstructure of dictionary articles. For the Frisian equivalents of the Dutch entries, it is important not only to show by way of examples, collocations, structures and expressions how the word can be used, but also, when necessary, to provide these equivalents with labels showing restrictions on usage. In many cases, the Dutch quotes in the dictionary are accompanied by more than one Frisian translation, varying from quite literal, but often stiff, Frisian to a more natural language use, in line with the concept of 'proscription'. The implication is that the ONFW is neither a purely descriptive nor a purely prescriptive dictionary. Proscription 'gives the lexicographer the opportunity to include different variants, but also to indicate the recommended form' (Bergenholz and Gouws 2010, 48). It means that the dictionary users can choose which kind of Frisian they wish to use (Van der Kuip and Visser 2018).

The ONFW has been set up in cooperation with the Dutch Language Institute in Leiden. The corpora of modern Dutch were taken from its monolingual *Algemeen Nederlands Woordenboek* (General Dutch Dictionary, ANW), with its dictionary writing system adjusted for the bilingual ONFW. The first online release is due in 2023 (http://frysker.nl) (see Duijff and Van der Kuip 2018).

2.2 Historical lexicography

In the 1950s, the Frisian Institute of the University of Groningen and the Fryske Akademy divided up the tasks in the field of Frisian Studies between them. Old and Middle Frisian (traditional periodization 1550-1800) would be described lexicographically in Groningen, Modern Frisian in Leeuwarden. In the 1990s, however, the lexicography of Old and Middle Frisian was in danger of not receiving proper attention at the university, and so this task was duly transferred to the Fryske Akademy.

2.2.1 Old Frisian lexicography

Linguists in the Netherlands began taking an interest in Old Frisian as far back as the seventeenth century (Franciscus Junius, Janus Vlitius). Later, and especially in the nineteenth century, interest grew mainly among German scholars. Two dictionaries of Old Frisian published in 1840 (Von Richthofen) and 1925 (Holthausen) do not include all extant Old Frisian. These dictionaries were supplemented in the early twenty-first century by an etymological dictionary, based on a manuscript that is assumed to represent Old Frisian in one of its ‘purest’ forms (Boutkan and Siebinga 2005), and an Old Frisian concise dictionary (Hofmann and Popkema 2008).

Several scholarly editions of Old Frisian manuscripts have been published in the course of time, with a glossary or a complete dictionary. Nowadays, the Fryske Akademy is working on a database of all Old Frisian texts, suitable for compiling an Old Frisian dictionary.

2.2.2 Middle Frisian lexicography

The lexicography of Middle Frisian began with a dictionary on the writings of the seventeenth-century poet Gysbert Japicx (Epkema 1824). This was followed by scholarly editions of Middle Frisian manuscripts and prints, among which can be found an explanatory edition of a large collection of idioms from 1614 (Van der Kuip 2003). In the early 1990s, the Fryske Akademy compiled a database of all extant Middle Frisian texts, which were subsequently lemmatized (POS-tagging). Even so, this period of Frisian has not been fully opened up lexicographically.

2.3 Miscellaneous

2.3.1 Dictionaries of dialects and mixed languages

The infrastructure of the ONFW is also used to compile a new and more extensive version of the *Stellingwarfs Woordeboek* (Bloemhof 1994, 1997, 2001, 2004). Stellingwarfs is a variety of Low Saxon spoken in the southeast of the province of Fryslân.

This brings us to the lexicography of peripheral Frisian dialects and Frisian-Dutch mixed languages. Recent examples of this are Duijff (1998) for several urban dialects, Visser and Dyk (2002) for the island dialect of Schiermonnikoog, Buwalda et al. (2013) for the mixed language of the municipality of Het Bildt, Oud and Dyk (2016) for the island dialect of Ameland, and Blom and Dyk (2019) for the dialect of the town of Hindeloopen. The metalanguage of these dictionaries is Dutch, but they also offer Frisian equivalents of the entries, to serve linguists. Most of these dialects and mixed languages already had smaller or more limited dictionaries, which served the 'dialectological information' of the WFT.

2.3.2 Bilingual dictionaries in which the second language is not Dutch

Frisian bilingual lexicography also involves languages other than Dutch. There is a Danish-Frisian/Frisian-Danish dictionary (Hoekema and Tams Jørgensen 1968) as well as a Frisian-English dictionary (Dykstra 2000) and, in a rather surprising combination, a Frisian-Japanese/Japanese-Frisian dictionary (Kodama 2004; 2015).

2.3.3 Terminology

One of the Fryske Akademy's other tasks in Frisian language development is language elaboration. At the end of the twentieth century, the Akademy published a lexicon of administrative language (De Haan 1989) and a legal dictionary (Duijff 2000). Both are intended for a broad group of civil servants and lawyers who are obliged to translate their professional Dutch jargon into Frisian in everyday practice. Another publication was a terminology list for non-language subjects in secondary education (De Haan 2000).

2.3.4 Taalweb Frysk

In 2015, the Fryske Akademy launched the online *Taalweb Frysk* (Frisian Language Web) (Dykstra, Duijff, Van der Kuip and Sijens 2015). This platform offers a standardized list of Frisian words, an online spelling checker, a machine translation tool and a dictionary portal. These applications give Frisian learners and native speakers a unique set of digital language tools that can help them to write Frisian and that is updated and improved regularly.[1]

3. Epilogue

For a minority language, Frisian has a rich variety of lexicographical output, as Popkema (2010, 83) shows, ranging from monolingual and bilingual dictionaries and lexicons to meta-lexicographical literature. We outlined this briefly in section 2 above.

The basic lexicographical facilities for Frisian are quite good, but this holds only for contemporary lexicography. As shown in section 2.1, there is a scholarly dictionary, the WFT, for Modern Frisian from 1800 onwards. With regard to earlier language periods, there is no Frisian counterpart of, for instance, the Middle Dutch Dictionary (MNW). Whether such counterparts will ever materialize remains to be seen. Historical lexicography is no longer as important to linguistics as it once was, and interest in Frisian studies appears to have shifted to synchronic linguistic research, with Frisian-Dutch bilingualism as an important component. For these reasons, the motto for the future of Frisian historical lexicography will have to be ‘cooperation’ (see Deutsch 2019 and Schoonheim 2019), i.e. cooperation with specialized historical linguists from—in the first place—Dutch studies (INT, Leiden) and German studies (IDS, Mannheim; Deutsches Rechtswörterbuch, Heidelberg). Clearly, there is still a great deal of work to be done.

Bibliography

Dictionaries

ANW = Algemeen Nederlands Woordenboek. (anw.ivdnt.org) [General Dutch Dictionary]

Bloemhof, H. 1994, 1997, 2001, 2004. Stellingwarfs Woordeboek. Oldeberkoop: Stichting Stellingwarver Schrieversronte. [Dictionary of the Dialect of Stellingwerf]

[1] The original website (taalweb.frl) with Frisian language tools was transformed to a new website (frysker.nl) in autumn 2021.

Blom, G., and S. Dyk. 2019. Graet Hylper Wordebook. Ljouwert/Leeuwarden: Afûk. [Dictionary of the Dialect of the Small Town of Hindeloopen]

Boutkan, D., and S.M. Siebinga. 2005. Old Frisian Etymological Dictionary. Leiden/Boston: Brill.

Buwalda, H.S., G. Meerburg, and Y. Poortinga. 1952. Frysk Wurdboek. Nederlânsk-Frysk. Bolswert: A.J. Osinga. [Frisian Dictionary. Dutch-Frisian]

Buwalda, H.S., G. Meerburg, and Y. Poortinga. 1956. Frysk Wurdboek. Frysk-Nederlânsk. Bolswert: A.J. Osinga. [Frisian Dictionary. Frisian-Dutch]

Buwalda, H.S., S.H. Buwalda, A.C.B. van der Burg, and H. Sijens. 2013. Woordeboek fan 't Bildts. Tweede ferbeterde en anfulde edisy. Sint Annaparochie: HDU. [Dictionary of the Dialect of the Bildt Region. Second improved and enlarged edition]

Dykstra, A. 2000. Frysk-Ingelsk wurdboek. Frisian-English Dictionary, with a Corresponding English-Frisian Word List. Ljouwert/ Leeuwarden: Fryske Akademy and Afûk.

Dykstra, A, P. Duijff, F. van der Kuip, and H. Sijens. 2015. Taalweb Frysk. Ljouwert: Fryske Akademy. (https://taalweb.frl). [Frisian Language Web]

Dijkstra, W. 1900, 1903, 1911. Friesch Woordenboek (Lexicon Frisicum). Leeuwarden: Meijer & Schaafsma. [Frisian Dictionary]

Duijff, P. 1998. Wurdlisten fan 'e Fryske stedsdialekten. Ljouwert/ Leeuwarden: Fryske Akademy. [Word Lists of the mixed Dutch-Frisian Urban Dialects]

Duijff, P. 2000. Juridisch Woordenboek Nederlands-Fries met een index Fries-Nederlands. Groningen: Martinus Nijhoff; Ljouwert/Leeuwarden: Fryske Akademy. [Dutch-Frisian Legal Dictionary, with a Frisian-Dutch Index]

Duijff, P., F. van der Kuip, R. de Haan, and H. Sijens. 2008. Frysk Hânwurdboek. Ljouwert: Fryske Akademy en Afûk. [Concise Frisian Dictionary]

DWB = Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm. 1854-1971. http://woerterbuchnetz.de/cgi-bin/WBNetz/wbgui_py?sigle=DWB [German Dictionary of Jacob Grimm and Wilhelm Grimm]

Epkema, E. 1824. Woordenboek op de gedichten en verdere geschriften van Gijsbert Japicx. Leeuwarden: Johannes Proost. [Dictionary on the Poems and Other Writings of Gysbert Japicx]

Haan, R. de. 1989. Wurdlist foar it offisjele ferkear. Ljouwert: Fryske Akademy. [Word List of Official Jargon]

Haan, R. de. 2000. Terminology fan de saakfakken. Ljouwert: Fryske Akademy. [Terminology List for Non-language Subjects]
Halbertsma, J.H. 1872. Lexicon Frisicum. A-Feer. Deventer: edited by Tjalling Halbertsma. (http://lexiconfrisicum.ivdnt.org)
Hoekema, T., and V. Tams Jørgensen. 1968. Deensk-Frysk wurdboek. Frisisk-Dansk ordbog. Groningen: Wolters-Noordhoff. [Danish-Frisian Dictionary. Frisian-Danish Dictionary]
Hofmann, D., and A.T. Popkema. 2008. Altfriesisches Handwörterbuch. Heidelberg: Universitätsverlag Winter. [Concise Old Frisian Dictionary]
Holthausen, F. 1925. Altfriesisches Wörterbuch. Heidelberg: Carl Winter's Universitätsbuchhandlung. [Old Frisian Dictionary]
Kodama, H. 2004. Frysk-Japansk wurdboek. Tokyo: Daigakusyorin. [Frisian-Japanese Dictionary]
Kodama, H. 2015. Japansk-Frysk wurdboek. Tokyo: Daigakusyorin. [Japanese-Frisian Dictionary]
MNW = Middelnederlandsch Woordenboek. 1885-1941. (gtb.ivdnt.org) [Middle Dutch Dictionary]
OED = Oxford English Dictionary. (https://www.oed.com)
Oud, A.G., and S. Dyk. 2016. Groat Amelander Woa'deboek. Ljouwert/ Leeuwarden: Fryske Akademy and Afûk. [Great Dictionary of the Dialect of the Island of Ameland]
Richthofen, K. von. 1840. Altfriesisches Wörterbuch. Göttingen: Dieterichsche Buchhandlung. [Old Frisian Dictionary]
Spoelstra, J., J. Post, and A. Hut. 2007. Prisma woordenboek Fries: Fries-Nederlands, Nederlands-Fries. Utrecht: Het Spectrum. [Concise Dictionary of Frisian: Frisian-Dutch, Dutch-Frisian]
Visser, W. 1985. Frysk Wurdboek. Nederlânsk-Frysk. Drachten / Ljouwert: A.J. Osinga Uitgeverij. [Frisian Dictionary. Dutch-Frisian]
Visser, W., and S. Dyk. 2002. Eilander Wezzenbúek. Woordenboek van het Schiermonnikoogs. Ljouwert/Leeuwarden: Fryske Akademy. [Dictionary of the Dialect of the Island of Schiermonnikoog]
WFT = Wurdboek fan de Fryske taal/Woordenboek der Friese taal. 1984-2011. Ljouwert/Leeuwarden: Fryske Akademy. (gtb.ivdnt.org) [Dictionary of the Frisian Language]
Winkler, J. 1898. Friesche Naamlijst (Onomasticon Frisicum). Leeuwarden: Meijer & Schaafsma. [List of Frisian Names]
WNT = Woordenboek der Nederlandsche Taal. 1882-2001. (gtb.ivdnt.org) [Dictionary of the Dutch Language]
Wumkes, G.A., and A. de Vries. 1918. Nederlandsch-Friesch Woordenboek. Snits/Sneek: A.J. Osinga. [Dutch-Frisian Dictionary]

Zantema, J.W. 1984. Frysk Wurdboek. Frysk-Nederlânsk. Drachten/Ljouwert: A.J. Osinga Uitgeverij. [Frisian Dictionary. Frisian-Dutch]

Other Sources

Bergenholz, H., and R.H. Gouws. 2010. "A Functional Approach to the Choice between Descriptive, Prescriptive and Proscriptive Lexicography." In: *Lexikos,* 20: 26-51.

Bosma. E., and N. Nota. 2019. "Cognate facilitation in Frisian-Dutch bilingual children's sentence reading: An eye-tracking study." *Journal of Experimental Child Psychology*, 189: 2-18.

Bremmer Jr, R.H. 1999. "De leksikografy fan it Aldfrysk." In: *In skiednis fan 'e Fryske taalkunde*, edited by A. Dykstra, and R.H. Bremmer Jr, 75-94. Ljouwert: Fryske Akademy. https://www.academia.edu/31068560/ In_skiednis_fan_e_Fryske_taalkunde [Lexicography of Old Frisian]

Bremmer Jr, R.H. 2001. "Lexicography of Old Frisian." In: *Handbuch des Friesischen/Handbook of Frisian Studies*. Edited by H.H. Munske, 653-657. Tübingen: Max Niemeyer Verlag.

Deutsch, A. 2019. "Old Frisian Terms in the 'Deutsches Rechtswörterbuch'—The Multilingual Approach of the Dictionary of Historical German Legal Terms." Lecture at the 10th International Conference on Historical Lexicography and Lexicology.
https://ivdnt.org/images/stories/onderzoek_en_onderwijs/publicaties/trefwoord/andreas-deutsch_oudfriese_termen.pdf

Dykstra, A. 1999. "De leksikografy fan it Nijfrysk." In: *In skiednis fan 'e Fryske taalkunde*, edited by A. Dykstra and R.H. Bremmer Jr, 191-212. Ljouwert: Fryske Akademy.
https://www.academia.edu/31068560/In_skiednis_fan_e_Fryske_taalkunde [Lexicography of Modern West Frisian]

Dykstra, A. 2001. "Lexicography of Modern West Frisian." In: *Handbuch des Friesischen/Handbook of Frisian Studies*, edited by H.H. Munske, 148-155. Tübingen: Max Niemeyer Verlag.

Dykstra, A. 2011. "J.H. Halbertsma als lexicograaf: studies over het Lexicon Frisicum (1872)" [dissertation, University of Amsterdam]. Ljouwert/Leeuwarden: Fryske Akademy/Afûk.
https://hdl.handle.net/11245/1.345550 [J.H. Halbertsma as a Lexicographer: Studies on the Lexicon Frisicum (1872)]

Duijff, P., and F. van der Kuip. 2018. "Lexicography in a Minority Language: A Multifunctional Online Dutch-Frisian Dictionary." In: *International Journal of Lexicography*, 31: 196-213.

Fryske Taalatlas 2020. [Atlas of the Frisian Language 2020] https://www.fryslan.frl/home/nieuws_3166/item/friezen-lezen-en-schrijven-meer-fries_22886.html

Haugen, E. 1966a. *Language Conflict and Language Planning: The Case of Modern Norwegian*. Cambridge (Mass.): Harvard University Press.

Haugen, E. 1966b. "Dialect, language, nation." In: *American Anthropologist*, 68: 922-935.

Hof, J.J. 1916. "In 'Nederlandsch-Friesch Woordenboek'." *Fryslân. Algemien Frysk wykblêd*. 1: 317. [A 'Dutch-Frisian Dictionary']

Kuip, F.J. van der. 2003. "De Burmania-sprekwurden. Santjinde-ieuske Fryske sprekwurden ferklearre en yn har tiid besjoen" [dissertation, Vrije Universiteit Amsterdam]. Ljouwert: Fryske Akademy. [The Burmania Proverbs. Seventeenth-century Frisian Proverbs Explained and Considered within the Context of Their Time]

Kuip, F.J. van der, and W. Visser. 2018. Introduction. In: *International Journal of Lexicography*. 31 (2): 127-131.

Popkema, A.T. 2010. "State of the Art of the Lexicography of European Lesser Used or Non-State Languages." In: Proceedings of the XIV Euralex International Congress., edited by A. Dykstra and T. Schoonheim, 65-98. Ljouwert/Leeuwarden: Afûk. https://euralex.org/wpcontent/themes/euralex/proceedings/Euralex%202010/000_Euralex_2010_03_Plenary_POPKEMA_State%20of%20the%20Art%20of%20the%20Lexicography%20of%20European%20Lesser%20Used%20or%20Non-State%20Languages.pdf

Schoonheim, T. 2019. "Old Frisian and Old Dutch, sibling languages in one dictionary?" Lecture at the 10th International Conference on Historical Lexicography and Lexicology. https://ivdnt.org/images/stories/onderzoek_en_onderwijs/publicaties/trefwoord/oudfries-oudnederlands-schoonheim.pdf

Sjölin, B. 1976. "'Min Frysk': een onderzoek naar het ontstaan van transfer en 'code-switching' in gesproken Fries." Groningen: Frysk Ynstitút, Rijksuniversiteit Groningen. ['Poor Frisian': An Investigation into the Origin of Transfer and Code-switching in spoken Frisian]

CONTRIBUTORS

Chiara Benati is Associate Professor of Germanic Philology at the University of Genoa, Italy. She has published essays on the Middle Low German influence of the phraseology of the oldest Swedish written sources, and on the Middle High German Dietrich epic and its reception in Scandinavia. Her current research interests include Middle High German literature, Middle Low German–Scandinavian language contact, and Faroese language and literature, as well as earliest (Low) German surgical treatises and their specialized terminology. She has published a wide range of articles, monographs and edited volumes in Italian, German and English. Examples of her work are *Storia delle lingue scandinave* (ECIG, 2011), *Dat Boek der Wundenartzstedye und der niederdeutsche chirurgische Fachwortschatz*, (Kümmerle Verlag, 2012) and "Painted Eyes, Magical Sieves and Carved Runes: Charms for Catching and Punishing Thieves in Medieval and Early Modern Germanic Tradition" In *Magic and Magicians in the Middle Ages and Early Modern Time*, edited by Albrecht Classen, 149-218 (de Gruyter, 2017). With Claudia Händl she edited *From Glosses to Dictionaries: The Beginnings of Lexicography* (Cambridge Scholars Publishing, 2019).

Chiara Bertulessi is Adjunct Professor of Chinese culture and postdoctoral researcher at the University of Milan, Italy. She also teaches Chinese at the University of Bergamo and LIMEC SSML (Milan) and is a specialist in lexicological and sociolinguistic characteristics of Chinese lexicography. Her publications include the monograph *La società cinese ieri e oggi: Testi, traduzioni e parole chiave* (Lumi, 2018) and several papers, such as "Sport e lessicografia nella Repubblica popolare cinese: analisi di un corpus di dizionari cinesi dedicati al lessico sportivo", *Sulla via del Catai* 21, 45-59 (2019).

Harald Bichlmeier is a senior researcher at the Sächsische Akademie der Wissenschaften in Leipzig, Germany, where he works on the etymological dictionary of Old High German. His publications include several monographs. His PhD was in the field of historical comparative syntax: *Ablativ, Lokativ und Instrumental im Jungavestischen – ein Beitrag zur altiranischen Kasussyntax*. (Baar-Verlag, 2011). With Güler Doğan

Averbek he has worked on versified dictionaries: *Almanca Tuhfe – Deutsches Geschenk (1916)* (Peter Lang, 2020), which is the topic of their article in this volume. He is also specialized in onomastics, and his *Studien zu nordostbayerischen Orts- und Flurnamen auf* -(n)itz *und* -(a)tz (Röll-Verlag) will appear in the course of 2021.

Andreas Deutsch is endowed professor at the law school of the Ruprecht-Karls-Universität Heidelberg and director of the Deutsches Rechtsworterbuch at the Heidelberger Akademie der Wissenschaften, Germany. He has a wide range of publications on German legal history and legal terminology and on civil law. Some of his recent publications include "Das Deutsche Rechtswörterbuch – ein Fachwörterbuch zwischen Recht, Sprache und Geschichte" In *Wortwelten: Lexikographie, Historische Semantik und Kulturwissenschaft*, edited by Volker Harm, Anja Lobenstein-Reichmann & Gerhard Diehl, 97-112, (De Gruyter, 2019), "Inszenierte Macht? Eine Zusammenschau der frühneuzeitlichen Illustrationen zur höchsten Reichsgerichtsbarkeit" In *Justiz und Verfahren im Wandel der Zeit. Gelehrte Literatur, gerichtliche Praxis und bildliche Symbolik*, edited by Eva Schumann, 133-184, (De Gruyter, 2017) and "Zur Symbiose zwischen "Zettelkasten" und "Datenbank" bei der Artikelerstellung im Deutschen Rechtswörterbuch" In *Historische Lexikographie zwischen Tradition und Innovation*, edited by Anja Lobenstein-Reichmann & Peter O. Müller, 271-286 (De Gruyter, 2016).

Güler Doğan Averbek is Associate Professor at Marmara University, Turkey. Her research area is Eastern languages and literatures, with special focus on Arabic and Persian and on Old Turkish literature. She has published numerous articles on these topics in Turkish. With Harald Bichlmeier she has worked on versified dictionaries: *Almanca Tuhfe – Deutsches Geschenk (1916)* (Peter Lang, 2020), which is the topic of their article in this volume. With Thoralf Hanstein she wrote "Oskar Rescher – Biographical Finds around Manuscripts, Books and Libraries" In *Collectors-Librarians-Readers: Contributions to the History of the Oriental Collections at the Berlin State Library*, edited by Sabine Mangold-Will, Christoph Raucha and Siegfried Schmitt, 387-449 (Vittorio Klostermann Verlag, 2021).

Fredric T. Dolezal is administrator for the *International Society for the History of Lexicology and Lexicography* and serves on the Advisory Board of the lexicographic journal *Lexikos*. He has research interests in history of linguistic semantics, philosophical grammars, and literary lexicography—especially as these linguistically-based studies raise questions concerning the dictionary as text and the literariness of literary theory. His most recent publications include a thematic issue of the *International Journal of Lexicography*, *Synonymy and Sameness of Meaning*, and the book chapters: “World Englishes and Lexicography” (*The Handbook of World Englishes*, Wiley Blackwell, 2020) and “Did Anne Maxwell Print John Wilkins's *Essay Towards a Real Character and a Philosophical Language* (1668)?” (Ward J. Risvold, co-author; *Historical Dictionaries in their Paratextual Contexts*, De Gruyter, 2018).

Maria Flaksman works as a postdoctoral research fellow at the Department of Indo-European Linguistics at Friedrich Schiller University Jena, Germany. She does research in iconicity studies, linguistic typology and historical linguistics and is a specialist of onomatopoeic words. Examples of this work are the papers “Iconic Treadmill Hypothesis. The Reasons Behind Continuous Onomatopoeic Coinage” In *Dimensions of Iconicity*, edited by M. Bauer, A. Zirker, O. Fischer and Ch. Ljungberg, 15-38, (John Benjamins, 2017) and “Pathways of De-Iconization: How Borrowing, Semantic Evolution, and Regular Sound Changes Obscure Iconicity” In *Operationalizing Iconicity*, edited by Pamela Perniss, Olga Fischer and Christina Ljungberg, 75-103, (John Benjamins, 2020). In her most recent work, she also includes psycholinguistic research methods, e.g. the co-authored “Iconicity and Second Language Visual Perception: A Psycholinguistic Study of English Imitative Words at Different De-Iconization Stages”, *Mathematics* 9(12), 1331 (2021).

Rachel Fletcher is a PhD student in English Language at the University of Glasgow, Scotland. She investigates the effects of periodization in English historical dictionaries, and traces the influence of these dictionaries on the wider field of Old and Middle English studies, addressing questions of how modern categorization and our reliance on the legacy of earlier scholarship have shaped our understanding of the past. She is the author of "“Most Active and Effectual Assistance” in the Correspondence of Sir William Dugdale and William Somner” (*Amsterdamer Beiträge zur älteren Germanistik 78*(2-3), 166-184 (2018).

Irena Fuková is a scientific collaborator at the Czech Language Institute in Prague, Czech Republic. She is specialized in historical lexicography and prepares a dissertation at the Charles University of Prague. She was one of the authors of the *Electronic Dictionary of Old Czech* and the lexical database of humanistic and baroque Czech. She has published several papers on morphological aspects of Old, Middle and New Czech.

Volker Harm is a senior researcher at the Zentrum für digitale Lexikographie der deutschen Sprache of the Akademie der Wissenschaften in Göttingen, Germany. He is the author of an introduction to lexicography *Einführung in die Lexikologie*. (Wissenschaftliche Buchgesellschaft, 2015). Volker Harm also edited several volumes on historical lexicology and lexicography (e.g. *Wortwelten. Lexikographie, Historische Semantik und Kulturwissenschaften*, De Gruyter, 2019) and on the history of German. He has also published numerous papers on these topics (e.g. "Between Science and Romanticism: The *Deutsches Wörterbuch* of the Brothers Grimm" In *The Whole World in a Book. Dictionaries in the Nineteenth Century*, edited by Sarah Ogilvie and Gabriella Safran, 73-92, (Oxford University Press, 2020).

Georgia Katsouda is a senior researcher at the Academy of Athens and teaches at the National and Kapodistrian University of Athens, Greece. She is specialized in dialectology, dialect lexicography, etymology and morphology. She is the author of several articles, books and edited volumes in Greek. Her book published in 2020, *The Dialect of Kythera. Description and Analysis*, (Society of Kytherian Studies) was originally published in Greek (Etaireia Kytheriakon Meleton, 2016). An example of her morphological work is "The Greek suffix -ozos: a case study in loan suffixation", *Journal of Greek Linguistics*, 232-265 (2016).

Abdusalom Mamadnazarov is a professor at the Tajik National University, Tajikistan. He is a leading specialist of Tajik lexicography and the author of the *Oxford English-Tajik Dictionary* (2018), also available online in Oxford Global Languages, and a monograph on the formation and development of Tajik translation lexicography (2016). He translated several books from English into Tajik and from Tajik to English.

Io Manolessou is a senior researcher at the Academy of Athens, Greece. She is specialized in historical linguistics, dialectology, lexicography and classical philology and has published more than fifty papers on these topics. She currently works on the compilation of the historical dictionary of modern Greek and on the Asia Minor dialects that are in danger of extinction. She has participated in a number of international research projects on Greek dialectology and linguistics and she is one of the authors of *The Cambridge Grammar of Medieval and Early Modern Greek*, (Cambridge University Press, 2019).

Katrin Menzel is a postdoc researcher at Saarland University, Germany. Her main research interests are historical linguistics, discourse analysis and pragmatics. Her dissertation (2017) is a study on ellipses as omission phenomena in English and German. With Stefania Degaetano-Ortlieb she studied the diachronic development of combining forms in English scientific texts over a time span of 350 years (*Lege Artis: Language yesterday, today, tomorrow*, 2017).

Ana Mihaljević is a faculty member at the University of Pula, Croatia. She obtained her PhD (2014) at the University of Zagreb, with a study of the syntactic characteristics of Croatian Glagolitic (the oldest Slavic script), based on texts translated from Latin. She continues to work on historical linguistics and language contact. With Jurica Polančec she published a paper on Croatian relativizers (*Jezikoslovlje* 2016: 429-451).

Markéta Pytlíková is a researcher at the Czech Language Institute in Prague, Czech Republic. She is specialized in diachronic lexicology and lexicography, diachronic textology with a special focus on Old Czech biblical texts and their translation and has published articles and collective volumes on these topics in Czech. She is the head of the department of language development and project manager of *Web Vocabulary,* an Internet application (launched in 2006) featuring text, image and sound resources for learning about historical Czech, along with tools for editing and analyzing historical texts.

Janneke Spoelstra is a lexicographer at Fryske Akademy in Leeuwarden, the Netherlands. She is co-author of a pocket translation dictionary of Frisian (*Prisma woordenboek Fries*) and is the author of a short and popular introduction to the Frisian language (*De kleine Fries voor Dummies*). She currently works on the Online Dutch-Frisian dictionary. She is also a creative writer in Frisian.

Petra Storjohann is a scientific collaborator at the Leibniz-Institut für Deutsche Sprache, Germany. She is the director of the Paronym dictionary and her main research interests are lexical semantics and lexicography. She has a wide range of publications (papers, monographs and collective volumes) in German and English on these topics. She has published papers in the Routledge Handbooks Series: "Lexicography" In *German lexicography in the Internet era*, 568-585 (2018) and "Sense relations" In *Semantics*, 248-265 (2015). She will be co-host of EURALEX 2022.

Hans Van de Velde is chair of sociolinguistics at Utrecht University and senior researcher at the Fryske Akademy in Leeuwarden, the Netherlands. He is a specialist in language variation and change, sociophonetics and standardization processes. At the Fryske Akademy he researches the Frisian minority language, Dutch and the mixed varieties spoken in the Friesland province (the Netherlands). He is responsible for the digital research infrastructure for Frisian and the development of Frisian language tools, such as the online Dutch-Frisian dictionary, spell checkers, automatic translation and speech recognition.

He co-edited several volumes *Language Variation–European Perspectives VIII* (with Nanna H. Hilton and Remco Knooihuizen, (John Benjamins, 2021) being the most recent one. He recently published papers on the use of Twitter data for the study of language change in low-resource languages (Dijkstra et al. 2021, *Frontiers in Artificial Intelligence*), the link between variation and perception and production in sound change (Pinget et al. 2020, *Language and Speech*), and on the formation of neologisms in Frisian (with Hindrik Sijens, 2020, *Dictionaries*). He organized the 10th International Conference on Historical Lexicography and Lexicology in Leeuwarden.

Frits van der Kuip was until his retirement in 2021 a linguist and lexicographer at Fryske Akademy in Leeuwarden, the Netherlands. He had a special interest in Frisian idioms and in historical and contemporary lexicography. He was one of the guest editors of a special issue of the *International Journal of Lexicography* on the *Lexicography of Smaller Languages: Between Foreignism and Purism* (2018, 31). With Eric Hoekstra he published an in-depth study of the Frisian loan (from Dutch) "pindakaas" ("Peanut butter, compositionality and transparency in loan translations"*, Taal en Tongval*, 137-156, (2018).

Willem Visser is a linguist and lexicographer at Fryske Akademy in Leeuwarden and an Assistant Professor at University of Groningen, the Netherlands. He currently works on the Online Dutch-Frisian dictionary. He is also the author of the part on Frisian phonology in Taalportaal, an online comprehensive and authoritative scientific grammar for Dutch, Frisian and Afrikaans (taalportaal.org). He was one of the guest editors of a special issues of the *International Journal of Lexicography* on the *Lexicography of Smaller Languages: Between Foreignism and Purism* (2018, 31). He was one of the organizers of the 10th International Conference on Historical Lexicography and Lexicology in Leeuwarden.

Kateřina Voleková is a scientific collaborator at the Czech Language Institute in Prague, Czech Republic. She is specialized in historical lexicography and lexicology, textology and Old Czech. She is one of the authors of Web Vocabulary and the *Electronic Dictionary of Old Czech* and is the coordinator of the Diabible project. She has published mainly in Czech, but recently she has co-authored papers in English on “Palaeoslovenisms in the Second Translation of the Old Czech Psalter” In *Old Church Slavonic Heritage in Slavonic and Other Languages*, edited by Ilona Janyšková et al., (2021) and papers in the journal *Museikon* (2019, 2020) on “The Musical Instruments in the Early Vernacular Translations of the Psalms”.