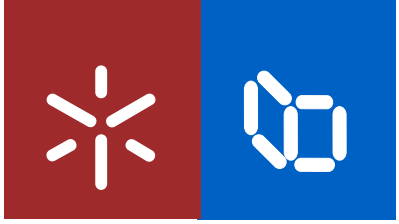
Universidade do Minho
Instituto de Letras e Ciências Humanas

Olha Vasylivna Morys

Bilingual onomasiological online dictionaries: the preparation of the English-Spanish Summer Camp Dictionary

July 2019

Bilingual onomasiological online dictionaries: the preparation of the English-Spanish Summer Camp Dictionary
Olha Vasylivna Morys
UMinho|2019



Universidade do Minho
Instituto de Letras e Ciências Humanas

Olha Vasylivna Morys

Bilingual onomasiological online dictionaries: the preparation of the English-Spanish Summer Camp Dictionary

Master Thesis
European Master in Lexicography

Supervised by
Professor Idalete Maria da Silva Dias
Professor Álvaro Iriarte Sanromán

July 2019

ACKNOWLEDGMENTS

There are many people who have contributed to the accomplishment of this dissertation and supported me throughout the whole course of study.

Firstly, I would like to express my sincere gratitude to my academic advisor, **Professor Idalete Maria da Silva Dias**, for supervising this thesis and continuous support in every step. This work would not be the same if not for her guidance, insightful and detailed comments, suggestions for improvement. I would also like to thank her for taking care of my accommodation issues and making my stay in Portugal as comfortable as possible.

I would also like to show my deepest gratitude to the committee, including, **Professor Álvaro Iriarte Sanromán**, **Professor Carlos Valcárcel Riveiro** and **Professor Idalete Maria da Silva Dias**.

My grateful thanks goes to the EMLex Coordinator, **Professor Stefan Schierholz** and all the **EMLex Consortium** for providing financial support and organizing the studying process in an interesting and efficient way. I will always carry positive memories about the EMLex.

I take this opportunity to express my profound gratitude to my parents **Mr. Vasyl Morys** and **Mrs. Svitlana Morys** for their unceasing encouragement. To my son **Dimitri** and husband **Gerardo** for coming with me to Portugal and motivating me during the last semesters of the studies.

Finally, I would like to say thanks to my friends and classmates of the EMLex program who have always helped me out to check the libraries of their home universities, advised new sources and shared their experiences.

I am grateful to the God for bringing all of these people to my life.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

ABSTRACT

Bilingual onomasiological online dictionaries: the preparation of the English-Spanish Summer Camp Dictionary

Bilingual onomasiological dictionaries generate a considerable interest because of two main reasons. On the one hand, thematic dictionary research is a less explored field of lexicography. On the other hand, bilingual onomasiological dictionaries combine several dictionary types. Since the Internet has established itself as a platform for online dictionaries, it is important to study how onomasiological dictionaries are organized on an electronic medium. The current thesis aims at analyzing two monolingual thematic online dictionaries, namely *BerufeNet* and *Computer Glossary*. Search options and navigation structures of these dictionaries receive special attention. The key focus of the dissertation is to create a prototype of the English-Spanish Summer Camp Dictionary (ESSCD), which involves the detailed description of the dictionary's sources, users and their needs, usage situations and functions. The methodology used to elaborate the ESSCD's prototype involves six main steps: (1) create a corpus; (2) extract keywords; (3) extract collocations; (4) generate concordances; (5) create an XML database; (6) design a web page. The current thesis discusses the above-mentioned steps as well as presents the data categories of an ESSCD article.

Keywords: electronic lexicography, lexicographical process, onomasiological dictionaries, specialized dictionaries

RESUMO

Dicionários onomasiológicos bilíngues on-line: a preparação do Dicionário de Acampamento de Verão Inglês-Espanhol

Dicionários onomasiológicos bilíngues geram um interesse considerável por duas razões principais. Por um lado, a pesquisa de dicionários temáticos é um campo pouco explorado da Lexicografia. Por outro, os dicionários onomasiológicos bilíngues combinam vários tipos de dicionários. Com a consolidação da Internet como uma plataforma para dicionários on-line, é importante observar como os dicionários onomasiológicos são organizados no meio eletrónico. A presente tese visa analisar dois dicionários on-line temáticos monolíngues, nomeadamente, o *BerufeNet* e o *Computer Glossary*. As opções de pesquisa e as estruturas de navegação desses dicionários recebem atenção especial. O foco principal da dissertação é criar um protótipo do Dicionário de Acampamento de Verão Inglês-Espanhol (ESSCD), que envolve a descrição detalhada das fontes do dicionário, utilizadores e suas necessidades, situações e funções de uso. A metodologia implementada na elaboração do protótipo do ESSCD envolve seis etapas principais: (1) criar um corpus; (2) extrair palavras-chave; (3) extrair colocações; (4) gerar concordâncias; (5) criar uma base de dados XML; (6) projetar uma página *web*. A presente tese analisa os passos acima mencionados, bem como apresenta as categorias de dados de um artigo ESSCD.

Palavras-chave: dicionários especializados, dicionários onomasiológicos, lexicografia eletrónica, processo lexicográfico

TABLE OF CONTENTS

Abbreviations

BNC: British National Corpus

BYU: Brigham Young University

COCA: Corpus of Contemporary American English

CSS: Cascading Style Sheets

Disco: European Dictionary of Skills and Competences

DoL: Dictionary of Lexicography

DTD: Document Type Definition

DWDS: Digitales Wörterbuch der deutschen Sprache

ED: Electronic Dictionary

ESSCD: English-Spanish Summer Camp Dictionary

GDEX: Good Dictionary Examples

IATE: Interactive Terminology for Europe

IENA: International Exchange of North America

IPA: International Phonetic Alphabet

KWIC: Keyword-in-Context

L1: Native Language

L2: Foreign Language

LSP: Language for Special Purposes

MI: Mutual Information

MLD: Monolingual Learner's Dictionary

PCDATA: Parsed Character Data

POS: part-of-speech

TEI: Text Encoding Initiative

XML: Extensible Markup Language

XSLT: Extensible Stylesheet Language Transformations

List of Figures

List of Tables

CHAPTER 1 INTRODUCTION

Since the beginning of the 21st century the Internet and innovative technologies have opened up new horizons for every branch of Humanities, including Lexicography. Carr's prediction that "the Internet will ultimately influence lexicography along with all fields of knowledge" (Carr, 1997, p. 221) has come true. The Internet has, indeed, established itself as a platform for online dictionaries. However, when it comes to *bilingual thematic dictionaries* in particular, metalexicographical discussion as well as the market offer appears to be rather limited.

Working as an application assistant for a summer camp staff placement agency, I have noticed, that the majority of program participants[1] were constantly struggling with filling in the application form. It seemed that the applicants lacked adequate support from reference works available online. That is how the idea of creating the *English-Spanish Summer Camp Dictionary* (henceforth ESSCD) was born. We decided to design a dictionary that would facilitate completion of application documents.

On the other hand, *onomasiological dictionaries* have hardly received any proper metalexicographic consideration until now, thus belonging to one of the most unexplored fields of Lexicography. Hartmann points out that the lexicographical research on *onomasiological dictionaries* is "immature, episodic and superficial" (Hartmann, 2005, p. 8). Nevertheless, we argue that *bilingual onomasiological online dictionaries* in particular are generating considerable interest in terms of their *hybrid genre* type[2]. Numerous researchers suggest that *thematic dictionaries* can be further developed in an electronic format (e.g. Jackson, 2002; Stark, 2011). We also maintain this point of view. As dictionaries are undergoing an online transformation, analysis of the structures of a thematic dictionary in online production form becomes fundamental.

In this chapter we clarify the focus of the thesis and specify its particular tasks and objectives. The introductory section defines the scope of the dissertation and introduces the previous research on thematic dictionaries. This chapter also includes definitions of the key terms used in the thesis. Last but not least, we present the structure of the thesis.

1.1 Scope of the Dissertation

This thesis deals with onomasiological online dictionaries, contributing to online thematic dictionary research. As far as we know, there are no bilingual onomasiological online dictionary products offered

[1] See Section 2.3 For whom is the ESSCD? User Profile.

[2] See Chapter 3 ESSCD as a Hybrid Lexicographic Genre.

on the market. Hence, we will focus on *monolingual thematic online dictionaries*. We do not aim at conducting an in-depth study, but rather at providing a general overview of the principal features of a thematic online dictionary. The thesis will not delve into the relatives of thematic dictionaries: thesauri and synonym dictionaries. Notwithstanding, we do not underestimate the importance of research into such reference works.

Since the main goal of the current Master thesis is to elaborate a prototype of a bilingual onomasiological online dictionary, we concentrated on the *lexicographic process* and its phases. Landau (2001) identifies three phases of the lexicographic process: *planning, writing and production*. In practice there are no fixed boundaries between the lexicographic phases especially when compiling an online dictionary. The lexicographic phases often run parallel to each other: "[c]harakteristisch für lexikographische Prozesse bei Internetwörterbüchern ist also ein zirkulärer, nicht-linearer Verlauf, an dem im Vergleich zu Printwörterbüchern nicht überwiegend Lexikographen, sondern auch Computerlinguisten und Informatiker beteiligt sind und in den auch die potenziellen Nutzer verstärkt einbezogen werden können und müssen [...]" (Hildenbrandt, V. & Klosa, A., 2016, p. 4). This tendency also applies to the ESSCD, as its planning, writing and production phases overlap: "working on a dictionary usually [...] means that while still writing the dictionary, new material may be added to the corpus, corrections for entries already published are gathered, headwords and cross-references will be supplemented, and modifications in the original concept will become necessary" (Klosa, 2013, p. 519).

Wiegand (1998) draws a distinction between the lexicographic process without computer use (*lexikographischer Prozess ohne Computereinsatz*) and the computer-assisted lexicographic process (*computerunterstützter lexikographischer Prozess*). The main objective of the computer-assisted lexicographic process is to create a dictionary adapted to the electronic medium. Wiegand determines 5 phases (see Figure 1) of the computer-assisted lexicographical process: (1) preparation, (2) data acquisition, (3) computerization, (4) data processing and (5) data evaluation[3]. Moreover, Klosa (2013) adds the sixth phase - preparation for online release, and speaks of a computer-lexicographical process. Fuertes-Olivera and Tarp (2014, p. 85) put forward three phases relevant for specialized online dictionaries: *pre-compilation, compilation and post-compilation.* This dissertation is based on the pre-compilation and compilation phases. The potential user group has been observed in a specific context and lexicographically relevant information needs were detected at the pre-compilation phase (see Sections 2.3, 2.4 and 2.5). The compilation phase embraces the dictionary concept (target audience, functions and type of the dictionary), preparation of the needed sources, elaboration and proofreading

[3] Appendix 1 illustrates the detailed tasks for each computer-assisted lexicographic process.

of test articles. The aspects of market research, staffing, calculations of expenses and benefits, lexicographer´s manual, financing, time plan with tasks and deadlines, online release preparation, archiving of dictionary data, dictionary version and its maintenance fall outside the scope of the thesis.

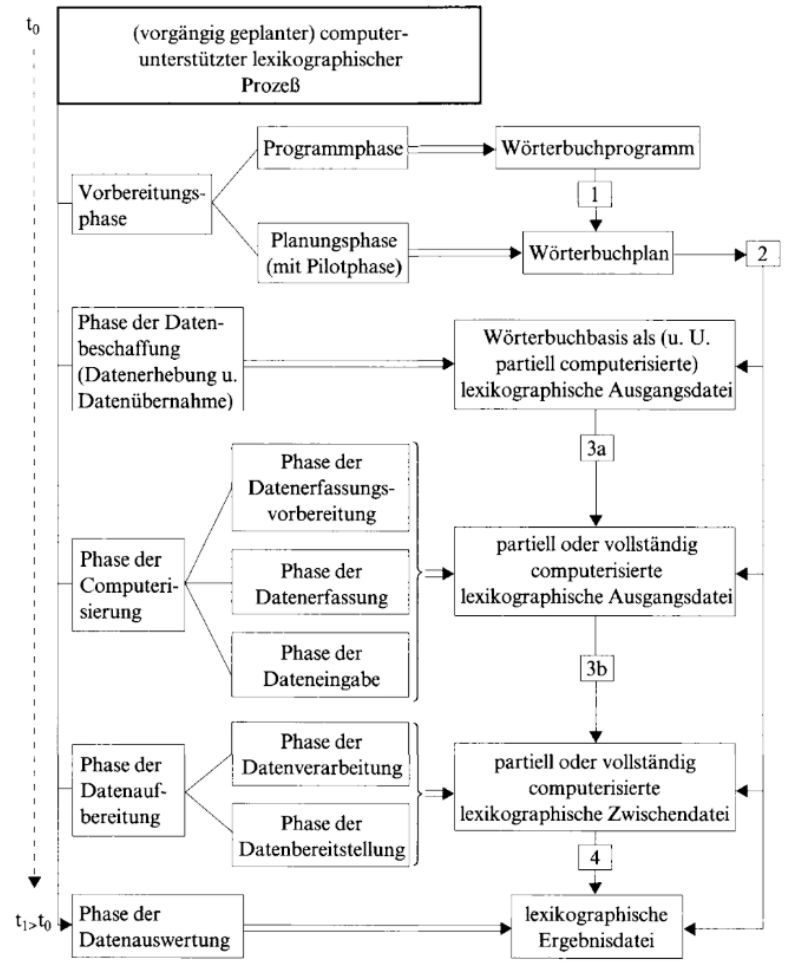


Abb. 1-41: Erste Veranschaulichung zum (vorgängig geplanten) computerunterstützten lexikographischen Prozeß; „——" bedeutet soviel wie *ist ein Subfeld von*; „⟹" bedeutet soviel wie *hat als Ergebnis*; „—1→" bedeutet soviel wie *legt den Rahmen fest für*; „—2→" bedeutet soviel wie *bestimmt die Ausführung von*; „—3a→" bedeutet soviel wie *wird (partiell oder vollständig) computerisiert zur*; „—3b→" bedeutet soviel wie *wird aufbereitet zu*; „—4→" bedeutet soviel wie *wird ausgewertet für*

Figure 1: Wiegand's visualization for the computer-assisted lexicographical process (Wiegand, 1998, p. 234)

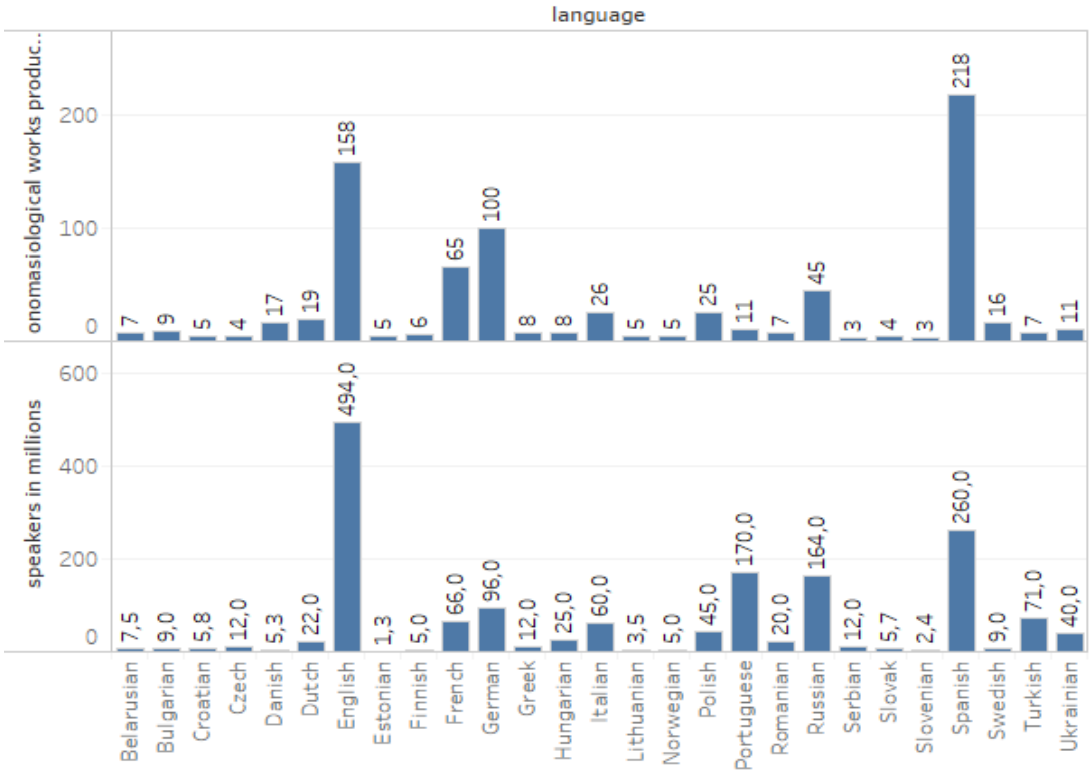
1.2 Objectives

The following are the principal objectives of the thesis:

- to provide an overall review of the access and cross-linking structures and data presentation within articles of thematic online dictionaries;
- to create a prototype of the ESSCD with a detailed description of the compilation process.

1.3 Significant Prior Research

This section attempts to briefly review the previous research on thematic dictionaries. Onomasiological dictionaries have hitherto been addressed only by a small number of researchers (Kipfer, 1986; Reichmann, 1990; Hüllen 1994; McArthur, 1986, 1998; Jackson, 2002; Hartmann, 2005; Stark, 2011 etc.). One of the most recent works on onomasiological dictionaries was carried out by Martin Stark in 2011. The researcher argues that bilingual thematic dictionaries are rather at the beginning of their development. In his work, these dictionaries are defined as a combination "of three lexicographic traditions: the bilingual, the thematic and the pedagogical" (Stark, 2011, p. 379). Stark observes that in the past onomasiological dictionaries were often the product of one author and mentions the *Longman Lexicon of Contemporary English* (1981) by McArthur and the *Oxford Learner's Wordfinder Dictionary* (1997) by Hugh Trappes Lomax as examples. Furthermore, the author discusses approaches to evaluating bilingual onomasiological dictionaries and provides recommendations for creating them. Although the survey on the usefulness of bilingual thematic printed dictionaries revealed that only 5% of learners of English and French as Foreign Languages use bilingual thematic dictionaries, 88% of informants find bilingual thematic dictionaries as useful as semasiological bilingual dictionaries. Interestingly, 65% of learners of English as a Foreign Language prefer bilingual onomasiological dictionaries to conventional bilingual dictionaries. The reason for such a result is the availability of notes and translation examples in L1. Onomasiological dictionaries have been addressed by Hartmann, who also conducted a survey and analyzed a number of onomasiological works published during the 20th century for about 20 languages. Hartmann has done tremendous work reviewing the *International Encyclopedia of Lexicography* and *Lexicographica Series Maior*, checking libraries and bookshops, involving more than fifty external experts from various countries to conduct an extensive survey on thematic dictionaries. The results of Hartmann's survey are visualized in Figure 2 and Appendices 2, 3. Surprisingly, languages with a lower number of speakers, such as Danish and Swedish have a much higher average of onomasiological works per million of speakers than Spanish or English. As can be seen from Figure 2, Spanish has the highest number of onomasiological dictionaries.



Sum of onomasiological works produced during the 20th century and sum of speakers in millions for each language. Details are shown for language. For pane Sum of onomasiological works produced during the 20th century: The marks are labeled by sum of onomasiological works produced during the 20th century. For pane Sum of speakers in millions: The marks are labeled by sum of speakers in millions.

Figure 2: Number of native speakers and onomasiological works per language. Based on Hartmann (2005). Visualized in Tableau Software

In terms of English onomasiological reference works, Hartmann highlights that "[t]here is no single publication dedicated entirely to the discussion of onomasiological dictionaries [...] the metalexicographic literature on this topic is often rather immature, episodic and superficial" (Hartmann, 2005, p. 8). Furthermore, Hartmann formulates desiderata for theoretical and practical lexicography. Particularly, the researcher calls for a more exhaustive bibliography of thematic dictionaries. Finally, Hartmann (2005, p. 15) concludes that "typological studies of onomasiological reference works other than the thesaurus genre are long overdue".

The chapter *Abandoning the alphabet* in Jackson's *Lexicography: An Introduction* discusses the drawbacks of an alphabetical listing. The researcher states that each word in A-Z arrangement is treated in isolation and "some words that belong together morphologically become separated" (Jackson, 2002, pp. 146-147).

Another significant researcher Tom McArthur describes onomasiological dictionaries from two perspectives: as a metalexicographer and as a lexicographer. What does this mean exactly? McArthur compiled the first thematic learner's dictionary, *The Longman Lexicon of Contemporary English* (1981),

and also provided broad perspectives, findings and results from the point of view of the actual dictionary-making process.

Nevertheless, all the above-mentioned research projects were related to printed dictionaries. Most scholars disregard thematic online dictionaries. When it comes to specialized online dictionaries, Fuertes-Olivera and Tarp (2014) have analyzed sixteen sources including *The Dictionary of Business and Management*, *The New Palgrave Dictionary of Economics Online*, *The Musikordbogen*, *The Glossary of Mortgage and Home Equity Terms*, *The Cambridge Business English Dictionary*, *TermFinder*, *The Business English-Spanish Glossary by A. D. Miles*, *The Glossary of FAO*[4] *Database and Information Systems (FAO Term Portal), Kicktionary*, *Interactive Terminology for Europe (IATE)*, *CercaTerm*, *The United Nations Multilingual Terminology Database (UNTERM)*, *The Multilingual Glossary*, *European Dictionary of Skills and Competences (Disco)*, *Genoma and EcoLexicon*. However, only two of these, namely *IATE* and *EcoLexicon*, seem to have the onomasiological access structure. Three of the sixteen analyzed dictionaries allow the user to select subfield, discipline or thematic area (*FAO Term Portal*, *TermFinder*, *CercaTerm*).

Some articles on onomasiological dictionaries were written by Sierra (2000, 2008) and Sierra & Hernández (2013). Notably, the author has elaborated and successfully tested the prototype of the *Diccionario Electrónico para la Búsqueda Onomasiológica*[5] with 33 terms of destructive phenomena. The dictionary is based on the following method: the user types keywords, and the system matches them by means of an inverted file in order to identify the possible term entries (Sierra, 2000). In addition, Sierra develops a method of automatic definition extraction for specialized onomasiological dictionaries (Sierra, 2008). Taking into consideration the previous research on thematic online dictionaries, we will attempt to discuss this dictionary type in this thesis.

We have taken the *Manual of Specialised Lexicography: The preparation of specialised dictionaries* (Bergenholtz & Tarp (Eds.), 1995) and *Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography* (Fuertes-Olivera & Tarp, 2014) as the main reference works for the thesis methodology. The theoretical frameworks of other lexicographers, Wiegand in particular, have been considered as well.

[4] Food and Agriculture Organization.

[5] Electronic Dictionary for Onomasiological Searching (translated by Sierra, 2000, p. 228).

1.4 Terms and Definitions

In this thesis the terms *onomasiological* and *thematic* are used interchangeably. *Candidates, participants* and *applicants* (of a summer camp program) refer to the target users of the ESSCD. *Text reception* and *text production* are used as synonyms for the terms *decoding* and *encoding* correspondingly. *Internet dictionary* is used as synonym for the term *online dictionary*.

Access structure "[d]ie Zugriffstruktur ist die Menge der Elemente und deren Ordnungsbeziehungen, auf die sich Benutzer während einer Zugriffshandlung stützen" (Engelberg & Lemnitzer, 2009, p. 272).

Collocations of a given word are statements of the habitual and customary places of that word (Firth, 1957, p. 181); Collocation is the cooccurrence of two or more words within a short space of each other in a text (Sinclair, 1991, p. 170).

Comparable corpora are "texts which, though composed independently in the respective language communities, have the same communicative function" (Laffling, 1992); "collection of similar documents that are collected according to a set of criteria, e.g. the same proportions of texts of the same genre in the same domain from the same period (McEnery and Xiao, 2007) in more than one language or variety of languages (Sinclair, 1996) that contain overlapping information (Munteanu and Marcu, 2005; Hewavitharana and Vogel, 2008)" (Skadiņa et al., 2012, p.7).

Corpus is "a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language" (Sinclair, 1996); "a collection of texts, of the written or spoken word, which is stored and processed on computer for the purpose of linguistic research" (Renouf, 1987, p. 1); "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research" (Sinclair, 2004).

Dictionary base is "die Menge der sprachlichen Dokumente und Wissensquellen, auf der die lexikographische Beschreibung des [...] Wörterbuchgegenstandes basiert" (Engelberg & Lemnitzer, 2009, p. 272).

Dictionary subject matter (Wörterbuchgegenstand) "ist die Sprache oder der Sprachausschnitt, die/den das [...] Wörterbuch aufgrund der vorhandenen [...] Wörterbuchbasis beschreibt." (Engelberg & Lemnitzer, 2009, p. 272).

Electronic dictionary "[t]he term electronic dictionary (or ED) can be used to refer to any reference material stored in electronic form that gives information about the spelling, meaning, or use of words. Thus a spell-checker in a word-processing program, a device that scans and translates printed words, a glossary for on-line teaching materials, or an electronic version of a respected hard-copy dictionary are

all EDs of a sort, characterised by the same system of storage and retrieval" (Nesi, 2000, p. 839); "[e]in e. W. [elektronisches Wörterbuch] ist ein Nachschlagewerk, das in digitalisierter Form auf einer CD-ROM, einer Diskette oder auf einem an das WWW angeschlossenen Server publiziert wird. Der Zugriff auf e. W.er ist nur mit Hilfe elektronischer Hilfsmittel möglich" (Engelberg & Lemnitzer, 2009, p. 271).

External subject classification is "a systematic arrangement of the subject field in question, delimiting this in relation to adjacent fields, with the purpose of identifying the material which is to form the empirical basis of the dictionary" (Bergenholtz & Tarp, 1995, p. 83).

Internal subject classification "establishes an overview of the subject area in question, thus forming the basis of the systematic structuring of the dictionary" (Bergenholtz & Tarp, 1995, p. 84).

Items are data-carrying entries from which the user can retrieve information relevant to fulfilling the genuine purpose of the specific dictionary. [...] e.g. items giving the pronunciation, morphology, part of speech, paraphrase of meaning, translation equivalents, illustrative examples, etc. (Gouws, 2014, p. 161).

Lemma selection „[m]it der Lemmaselektion werden diejenigen sprachlichen Einheiten identifiziert und ausgewählt, die Gegenstand der lexikographischen Beschreibungen sein sollen" (Engelberg & Lemnitzer, 2009, p. 246).

Lexicographical process „[e]in abgeschlossener lexikographischer Prozeß ist zu verstehen als die Menge derjenigen Tätigkeiten von Lexikographen, die ausgeführt wurden, damit ein bestimmtes Wörterbuch entsteht [...] Kalkulierbarkeit, Zerlegbarkeit, Kontrollierbarkeit, Reglementierbarkeit, Lehrbarkeit und Prüfbarkeit sind zentrale Eigenschaften, die lexikographische Herstellungsprozesse und die gesamte lexikographische Praxis mit Herstellungsprozessen anderer Art teilen" (Wiegand, 1989a, pp. 250-251).

Macrostructure "[u]nter der MAKROSTRUKTUR eines Wörterbuches verstehen wir die geordnete Menge seiner Lemmata" (Kunze & Lemnitzer, 2007, p. 79)[6].

Terminological classification is "a systematic listing of the LSP terminology of the subject field in question, for the purpose of ensuring that all LSP terms are captured" (Bergenholtz & Tarp, 1995, p. 84).

[6] Emphasis in original.

1.5 Structure of the Thesis

This dissertation is divided into five chapters. Following the introductory part, the thesis deals with the applied methodology in the second chapter: potential users and their information needs have been identified, usage situations and dictionary functions have been defined. Moreover, the second chapter describes material collection and data acquisition. The third chapter analyzes the ESSCD dictionary type and highlights the user research and user participation in online dictionaries. The discussion on online versus printed dictionaries gets special attention. Current online onomasiological dictionaries are considered in more detail. The fourth chapter focuses on the process of creation of an *XML* (Extensible Markup Language) database and the ESSCD structures. ESSCD's articles and their data categories are presented in the fifth chapter. Finally, we critically analyze the limitations and suggest some insights into future work, followed by conclusions.

CHAPTER 2 METHODOLOGY

> I certainly do not know all lexicographic projects past and present; but one of those I know not a single one was finished in the time and for the money originally planned.
>
> Zgusta, 1971, p. 348

During the lexicographical process "different decisions must be taken, actions must be done and different methods must be used" (Schierholz, 2015, p. 328). This statement also applies to the ESSCD. A few questions arise in the process of dictionary creation: what will the base of the dictionary be; how will the lemma candidate list be selected; what data categories should the dictionary include; which examples should be taken into the dictionary; how should the dictionary database be created; how should the microstructure, macrostructure and access structure be organized.

All in all, the lexicographical process of the ESSCD splits into the following six work steps: (1) corpus creation, (2) extraction of keywords, (3) extraction of collocations, (4) generation of concordances, (5) database creation, and (6) website design. A range of tools is used for each step (see Figure 3).

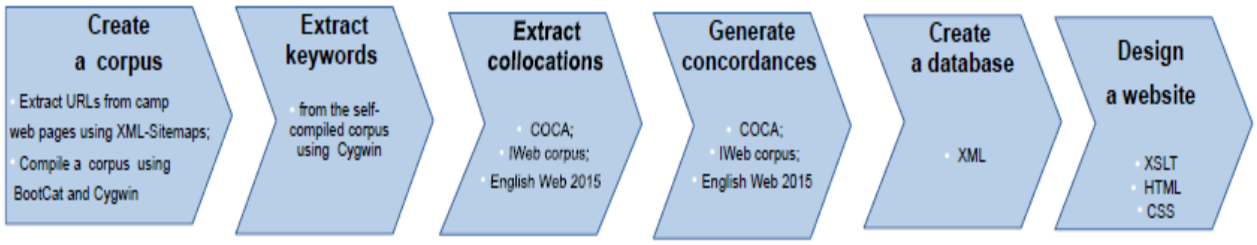


Figure 3: Steps and working tools for the compilation of the ESSCD

As can be seen from Figure 3, all the corpora used for the ESSCD will be searched with *concordances*. Concordances as a data extraction mode are commonly used in lexicographical practice. They extract phrases, where a certain headword occurs, and present it in the form of a list. To search corpora for the ESSCD we use the *Keyword in Context* (*KWIC*) concordance type.

As has already been mentioned above, the ESSCD database is developed using *Extensible Markup Language* (hereafter *XML*), which allows for the creation of the project specific tags. *Extensible Stylesheet Language Transformations* (henceforth *XSLT*) and *Cascading Style Sheets* (hereafter *CSS*) are likely to be employed for the dictionary web page design.

2.1 The Dictionary Basis

Regarding the dictionary base for specialized dictionaries, Bergenholtz and Tarp (1995, p. 90) speak about *empirical basis* which consists of three basic types: (1) *introspection*, (2) *existing literature*, (3) *texts.* Introspection refers to the lexicographer's own competencies. For the ESSCD, *introspection*

means first of all the lexicographer's cultural knowledge of summer camps and linguistic competence in both source and target languages. Given that the ESSCD is a bilingual dictionary, *introspection* includes the lexicographer's translation skills as well. Obviously, lexicographers cannot rely solely on their own skills. Hence, *existing literature* and *texts* are of paramount importance. For the preparation of the ESSCD, the following existing literature sources will be used: *Linguee English-Spanish Dictionary* for translation equivalents, web site *Indeed* for notes, grammar books of the English and Spanish languages for grammatical information. Texts for the ESSCD are corpora, which are described in the follow-up section of the thesis.

2.1.1 Primary Sources

Wiegand (1998, p. 140) describes the primary, secondary and tertiary sources as follows: „[d]ie *primären Quellen* sind vor allem (aber nicht nur) Texte, welche aus natürlichen oder quasi-natürlichen Kommunikationssituationen stammen, oder größere zusammenhängende Ausschnitte aus solchen. [...] Die Menge der primären Quellen, welche Texte oder größere zusammenhängende Ausschnitte sind, bildet das *lexikographische Korpus* [...]. Zu den *sekundären Quellen* gehören alle Wörterbücher, die nach dem Instruktionsbuch entweder obligatorisch oder fakultativ konsultiert werden sollen, und zu den *tertiären Quellen* gehören alle sonstigen Sprachmaterialien, die benutzt werden wie z.B. linguistische Monographien und Grammatiken [...]". In the frame of the current project we use general language corpora (COCA, iWeb Corpus, English Web Corpus 2015) and a self-compiled comparable corpus as primary sources. Additional search engines, for example *Indeed*, and application forms support lemma selection and writing of the dictionary articles. Bilingual dictionaries, for example *Linguee English-Spanish Dictionary*, and grammar books belong to the secondary sources of the ESSCD.

2.1.1.1 Corpora

Corpus evidence allows observing the behaviour of a word in its context, thus opening new perspectives for Lexicography. Several researchers have proposed definitions for the concept *corpus* (Atkins, Clear & Ostler, 1992; Sinclair, 1996; Pearson, 1998). In general, all the linguists agree that the corpus is a collection of authentic texts sampled in a systematic way which serves to represent a specific language. The corpus size is usually associated with its representativeness. The corpus is representative when "the lexical density does not alter when more texts are added" (Losey-León, 2015, p. 298). Sinclair (1996) defines three main characteristic of the corpus: *quantity, quality* and *simplicity*. The researcher determines a default value for each characteristic: *large* for quantity, *authentic* for quality and *plain text* for simplicity. In addition, Sinclair distinguishes between *text corpus* (or *whole text corpus*) and a

samples corpus. The latter refers to specialized corpora, "which do not contribute to a description of the ordinary language, either because they contain a high proportion of unusual features, or their origins are not reliable as records of people behaving normally" (Sinclair, 1996).

It goes without saying that corpora make it possible to identify the features of lexical units and to discover new collocations providing authentic examples of language in use. However, Geyken and Lemnitzer (2016, pp. 203-204) consider some significant limitations with respect to corpora as a dictionary base: "kein Korpus, egal welcher Größe kann eine lebende Sprache als Ganzes abbilden oder repräsentieren". The editors of DWDS note that corpora often consist of newspaper texts solely. Only limited specialized corpora are likely to contain transcripts of spoken language. Moreover, the analysis of a huge amount of data can also lead to errors, because lexicographers tend to make subjective decisions (Geyken & Lemnitzer, 2016).

2.1.1.1.1 Collecting the Comparable Summer Camp Corpus

This section deals with the compilation process of the *English-Spanish Comparable Corpus* and provides a comprehensive description of the steps followed in the corpus creation process. Creating a corpus for specialized dictionaries has its own requirements (Bergenholtz & Tarp, 1995, p. 94):

1. the corpus should cover all sub-fields of the subject matter in question;
2. the text types to be considered in the dictionary should be included in relation to their presumed relevance for the intended dictionary users and use situations.

The English-Spanish Comparable Corpus is a web-based corpus. Firstly, the sources were selected: 20 web pages of Canadian and American summer camps, American Camp Association page, Canadian Camping Association and a Canadian placement agency pages for English and 18 Spanish and Mexican summer camp pages, camp searcher page and the Mexican Association of Camps for Spanish. It should be noted that the content of web pages for summer camps in Spain differ slightly from the classical organization of a summer camp. In Spain summer camps either focus on learning English, i. e. language camps such as Campamento Náutico Bilingüe, Irish Summer Camp or they are dedicated to a particular activity: Rockcamp dedicated to music, Campus Experience to football, Club Hipico Miracampos to horse riding. The main question which every lexicographer asks before compiling a corpus deals with the size of the corpus. Albeit the exact number of words to achieve a representative corpus has not been established yet, several suggestions have been made so far. For a specialized corpus it is generally agreed that "corpus size depends on the corpus aim and no minimum or

maximum extension is particularly required" (Losey-León, 2015, p. 296). One million words is the size usually recommended for a specialized corpus (Bergenholtz & Tarp, 1995; Pearson, 1998).

Basically, there are two main ways of compiling a corpus from the Web: (a) *introducing a set of seed words to automatically search for pages with these words*, or (b) *crawling a number of web pages*. The first approach results in an enormous corpus with a lot of noise, which is time-consuming to edit and analyze. Hence, we determined the above-mentioned web pages[7] to gain a set of links and to convert them into plain text. *XML-Sitemaps*[8] and *Extract Links* provided all the links inside of the web pages. It was observed that *XML-Sitemaps* outputs a higher number of web pages than *Extract Links*. Moreover, *Extract Links* did not output any URLs on several occasions. Another disadvantage of *Extract Links* may be that the *BootCat*[9] often will not accept the extracted URLs because the output files are not well-formed. As we have noticed, *Extract Links* retrieves e-mails and phone numbers as URLs that are not processed by BootCat's "Custom URLs" option.

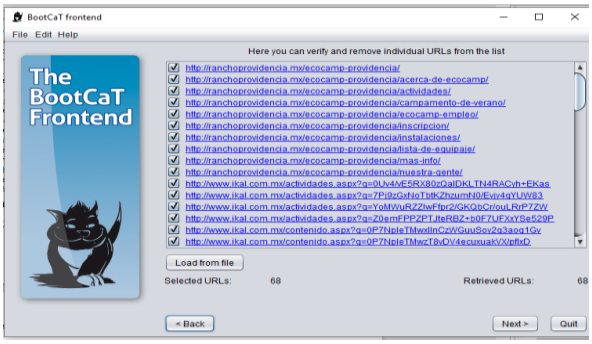


Figure 4: Corpus building from the URLs in BootCaT

Generally, summer camp websites do not include external web pages. However, they usually keep social media and foundation links, which should be deleted before corpus compilation. *BootCat* provides a separate .txt file for each URL, therefore the next step was to combine all the txt files to a single document using *Cygwin* command[10] *"cat *txt > corpus.txt"*, which allows numerous .txt files to be merged. In the text processing stage with *UDPipe*[11] we encountered problems that resulted from the fact that the corpus was too large and *UDPipe* was unable to generate an output file. Hence, the corpus was processed by parts. *UDPipe* carries out three tasks: *lemmatisation*, *tagging* and *parsing*. In terms of lemmatisation, it relates inflected forms of a word to its lemma. Tagging is also called part-of-speech

[7] The list of selected web pages can be found in the References section.

[8] *XML-Sitemaps* and *Extract Links* are tools, which quickly scrape internal and external links from web pages.

[9] *BootCat* is usually used for the seed words method. In the context of the current project own URLs were introduced (see Figure 4).

[10] The command was provided by Prof. Dr. Marcos García González, University of A Coruña. *Cygwin* is a collection of Open Source tools which allows the reuse of its source codes for creation of own programs (Cygwin FAQ. Available at https://www.cygwin.com/faq/faq.html).

[11] *UDPipe* is a text processing tool which tokenizes, tags, lemmatizes and parses corpora.

labelling, so *UDPipe* assigns a word class to each word and enables syntactic extractions. When the tool parses the corpus, it assigns a syntactic tree to every sentence. Once the corpus is compiled, we can move to its description. The Table below specifies the characteristics of the self-compiled comparable corpus. It is a web-based synchronic corpus in the field of summer camps.

Topic variety	Field of summer camps
Chronology	English-Spanish Comparable Corpus captures the most recent texts, web pages are continuously updated
Origin	As far as geographical origin is concerned, the self-compiled corpus embraces samples of the American and Canadian varieties; Spanish of Spain and Mexico
Mode	The corpus is focused on the written mode of the language

Table 1: Description of the self-compiled corpus

The statistical analysis of the English-Spanish Comparable Corpus was carried out in Sketch Engine. We consider that 3131316 words suffice for both the subject domain and the project goal, where the self-compiled corpus serves to complement the lemma candidate list. Small domain specific corpora are more likely to contain relevant concordances than huge general corpora. The major drawback of using small corpora is that the usage of some patterns, especially uncommon ones, is not represented. As a consequence the lexicographer lacks lexical evidence (Bernardini & Ferraresi, 2013). Hence, larger corpora, namely *Corpus of Contemporary American English* (COCA), *Iweb* and *English Web Corpus* (enTenTen15)[12], have been browsed to compile the ESSCD.

Being a comparable corpus, the self-compiled corpus consists of two subcorpora, English and Spanish (see Table 2). *Comparable* in this respect means that the selected texts belong to the same domain, but are not direct translations of each other. However, direct equivalents of phrases or terms tend to occur in both language versions/texts. Richness, availability and diversity are main advantages of comparable corpora in comparison with parallel ones (Skadiņa et al., 2012).

[12] See enTenTen: Corpus of the English Web. Available at https://www.sketchengine.eu/ententen-english-corpus/

	English subcorpus	Spanish subcorpus
Tokens	3 159 320	727 998
Words	2 500 100	631 216
Sentences	124 983	29 186
Lemmata	42 181	17 140

Table 2: Statistical information about the self-compiled corpus

After the texts were collected in an electronic format in both the English and Spanish languages, the self-compiled corpus was processed. Candidate term pairs were generated for each language by running the script *patterns.py.* The script outputs examples containing part-of-speech (POS) tagging and computes the statistical association of words in the text. It identifies instances according to three patterns: NOUN+NOUN, ADJ+NOUN and NOUN+ADP[13]+NOUN. It should be noted that in order to extract keywords and terms in Cygwin, some basic *Python* commands need to be employed, for example, to open the folder, to print the current folder, to list the folder's content etc. The next task consisted of a manual verification of the term pair candidates.

To sum up, the English-Spanish Comparable Corpus is a bilingual synchronic corpus of written texts, designed to represent the summer camp domain. One of the strengths of a web-based corpus is the accessibility of large quantities of text. On the other hand, compiling a corpus is a time-consuming and technically complex process. The self-compiled corpus was used to extract keywords and terms and to identify their translation equivalents. Bergenholtz and Tarp (1995) suggest that terms in a specialized corpus have to be identified for each language in order to detect the culture-dependent or language-dependent terms.

2.1.1.1.2 Corpus of Contemporary American English

COCA is a dynamic or a so-called monitor corpus, which is constantly being updated with new texts. It includes more than 560 million words from various genres starting from spoken language and fiction to academic journals and newspapers. Interestingly, the genre variety is balanced from one period to another. The search in COCA can be carried out by words, phrases, lemmata, wildcards or more complex searches. The search for collocates can be delimited by frequency and mutual information score. Mutual information (MI) is a measure of collocational strength. The default score in COCA is three, which is defined to avoid noisy examples. The higher the MI score, the stronger the association

[13] Adposition, a cover term for prepositions and postpositions, for example *in*, *to*, *during*.

between the words x and y is[14]. Besides, there is a possibility to search collocates by subgenres and to determine the distance of collocates[15]. COCA corpus provides metadata for all its texts. While browsing concordances, it is possible to consult the author, date, title and further publication information. Apart from that, COCA enables the user to select a token range to the left or right of the headword.

2.1.1.1.3 iWeb Corpus

IWeb is a 14 billion word corpus, i.e. 25 times bigger than COCA. iWeb belongs to BYU (Brigham Young University) corpora together with COCA, and offers the same search options. The distinctive feature of the iWeb Corpus is the possibility to create a separate virtual corpus and search it as a stand-alone corpus. IWeb has a more advanced architecture which enables, for instance, NOUN+NOUN searches. This feature is crucial for the ESSCD as many of its headwords are compounds (e.g. child care, office worker, arts and crafts etc.). For the top 60 000 words, the iWeb Corpus provides definitions, translations, pronunciation, images and videos, synonyms and collocates, hyponyms and hyperonyms, related words, concordances and clusters[16]. The *i* in the designation iWeb Corpus stands for[17]:

- immense: iWeb is large;
- insightful: iWeb allows to access specific topics such as nuclear energy;
- instantaneous: iWeb is fast;
- informative: as mentioned earlier, iWeb provides exhaustive information on 60 000 lemmas;
- integrated: the user can easily switch between words.

If one wants to access the text metadata while searching concordances or would like to search for more context, the iWeb Corpus redirects the user to the web page link. The shortcoming here is that some pages cannot be found. The result of the search is an error because sources have been removed, had their name changed, or are temporarily unavailable. As a consequence, it is impossible to see the full sentence or the authorship details in some cases. One more drawback of the iWeb Corpus is that the search and decision-making processes are time consuming. Working with this corpus may become a challenging task for the lexicographer who needs to review a long list of concordances to find a suitable example.

[14] Appendix 4 informs about how the MI score in COCA is computed.

[15] Corpus of Contemporary American English. Retrieved from https://www.english-corpora.org/coca/

[16] The iWeb corpus. Retrieved from https://corpus.byu.edu/iweb/help/iweb_overview.pdf

[17] Why the name iWeb? Retrieved from https://corpus.byu.edu/iweb/help/iweb_name.asp

2.1.1.1.4 English Web Corpus 2015

The English Web Corpus 2015 (enTenTen15) with around 15 billion words was created from the Web for Sketch Engine. Like many Sketch Engine corpora enTenTen may be browsed for word sketches, i.e. collocations, synonyms, keywords and terms and examples in context. The innovative feature of Sketch Engine is *good dictionary examples* (*GDEX*), which automatically finds good candidate sentences. Sketch Engine implements a good dictionary examples algorithm, which scores sentences according to certain parameters such as length of the sentence, occurrence of pronouns and proper names. This feature of the Sketch Engine assists lexicographers in faster selection of illustrative examples (Kilgarriff, 2013).

2.1.1.2 Other Primary Sources and Lemma Selection

2.1.1.2.1 International Exchange of North America Application Forms

To participate in a summer camp program, the applicants need to fill in an application form. International Exchange of North America (IENA) will be used as a source for the ESSCD lemmata list. IENA is one of J-1 visa sponsors[18], i.e it provides required documents for candidates to obtain a J-1 visa. The application form for a summer camp program contains the same fields for all visa sponsors. Therefore, any participant can consult the ESSCD regardless of the sponsor they apply with. The IENA application form constitutes the main source for lemma candidate list selection: "[q]uality in practical lexicography includes meticulous, goal-oriented selection of lemmata" (Bergenholtz & Tarp, 1995, p. 98). The IENA application consists of ten steps: personal information, education and background, job preference, work experience, skills, references, documents, medical history, visa information and emergency contact. The steps differ from each other in terms of length and the participants' background. The visa information is one of the shortest sections if the applicants apply for the first time, as in this case they do not have any visa history. The task of lemma selection is to identify the headwords which may be unfamiliar for the program participants. As the program applicants are required to have at least level A2 in English, not all the words should be taken into the ESSCD. In the case of the step "Visa Information", the terms *denial, visa type, sponsor, J-1 visa, legal residence* should be included.

[18] The full list of visa sponsors can be accessed on the web page of the US Department of State.

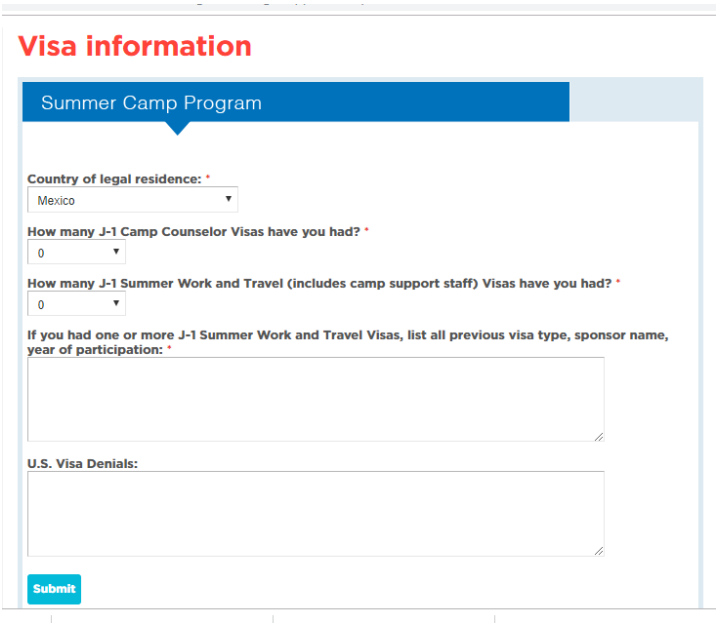


Figure 5: Visa information step in IENA's application form

The most significant section of the whole application form is the *Skills* section. Camp directors pay special attention to this section, when hiring candidates. For this reason the current thesis focuses on the lemma selection for the *Skills* section. It goes without saying that the lemma candidate list is open and new headwords can enter the list at all stages of the lexicographical process. The IENA application form also contributes to systematic *classification of the subject area*. Within this classification an *external subject field classification*, *an internal subject classification* and *terminological classification*[19] can be determined (see Bergenholtz & Tarp, 1995). *External subject field classification* delimits the field/fields to be covered. The external subject field classification for the ESSCD corresponds to the following 10 steps of the application form:

1. Personal information
2. Education and background
3. Job preference
4. Work experience
5. Skills
6. References
7. Documents
8. Medical history
9. Visa information

[19] The terms *external subject field classification*, *internal subject classification* and *terminological classification* are defined in the Section 1.4 Terms and Definitions.

10. Emergency contact

An internal subject classification functions as a drop-down menu of the *external subject field classification*, and is aimed "to ensure systematic representation of the subject field in the dictionary" (Bergenholtz & Tarp, 1995, p. 85). For the ESSCD it would look as follows:

1. Personal information

1.1. Basic info

1.2. Dates of availability

1.3. Permanent address

1.4. Mailing address

2. Education and background

2.1. Education

2.2. Dates of summer holiday

2.3. Background

3. Job preference

4. Work experience

4.1. Experience working with children

4.2. Employment and volunteer history

5. Skills

5.1. Camp Counselor

5.2. Support Staff

6. References

6.1. Reference

6.2. Applicant information

6.3. Rate the following

6.4. Overall

6.5. Reference signature

7. Documents

7.1. Health history form

7.2. Passport

7.3. Photo album

7.4. Photograph

7.5. Police background check

7.6. Proof of student status

7.7. Camp contract

8. Medical history

8.1. Medical history

8.2. Other background info

9. Visa information

10. Emergency contact

10.1. Contact address

10.2. Next of Kin's address

One may wonder why *dates of summer holidays* is a separate subsection. In fact, the accuracy of the holiday duration is especially important, because if the dates the applicant has selected in the application form do not correspond to the dates on the *proof of student status,* two options are possible (1) the visa sponsor will not send documents required for the J-1 visa; (2) in case that the sponsor has sent the documents, there is a high risk of visa rejection.

Terminological classification lists the *Language for Special Purposes* (LSP) terms related to the subject field of summer camps. Such listing ensures that the main vocabulary of the ESSCD is lemmatized. Research suggests that not only lemmata, but also non-lemmatic units (*nichtlemmatische Einheiten*) such as collocations, phrasal verbs, compounds and fixed expressions should be taken into the candidate list (Engelberg & Lemnitzer, 2009, p. 246). Dictionaries aimed at text production should include LSP terms and other non-common-language expressions according to Bergenholtz and Tarp (1995, p. 103). As can be seen from Table 3, the dictionary will lemmatize not only words but also phrases, i.e “a particular kind of lexeme which in itself consists of several lexemes or several grammatical words” (Bergenholtz & Tarp, 1995, p. 100).

5. Skills												
5.1. Camp Counselor								5.2. Support Staff				
5.1.1. Adventure	*5.1.2. Aquatics and Waterfront*	*5.1.3. Arts and Crafts*	*5.1.4 Horse Riding*	*5.1.5. Music*	*5.1.6. Performing Arts*	*5.1.7. Special Skills*	5.1.8. *Sports*	*5.2.1. Administration*	*5.2.2. Housekeeping*	*5.2.3. Kitchen*	*5.2.4. Maintenance*	*5.2.5. Security*
camping	diving	arts and crafts	stable management	bagpipes	ballet	animal care	aerobics	office manager	cleaning	baking	carpentry	security guard
climbing	fishing	camera work		drumming	cheer-leading	camp fires	archery	office worker	housekeeping	cook	electrical	
high ropes	sailing	candle making		Guitar	choreography	Chess	basketball		laundry	dish washer	gardening	
hiking	surfing	ceramics		Harp	circus skills	child care	board diving			food preparation	lawn mowing	
low ropes	swimming	costume design		Music	dance	Cookery	boat driving			general kitchen	maintenance	
	windsurfing	fine arts		Piano	hip-hop	digital editing	cross country			head chef	plumbing	
	canoeing	glass blowing		singing	lighting	Dressage	cycling			Waiter		
	jet skiing	jewellery making		violin	performing arts	Ecology	dodgeball					
	kayaking	knitting			set design	I.T.	equestrian					
		leather craft			sound technician	Leadership	fencing					
		loom weaving			stage management	Mechanic	field hockey					
		magic			tap dance	Nature	fitness					
		make up artist			theater	Newspaper	go-carts					
		mask making				Nutrition	golf					

media-art	print making	gymnastics	
metalsmith	Radio	lacrosse	
model making	recording studio	land sports	
painting	Robotics	life guarding	
photography	Rocketry	martial arts	
sewing	Science	motocross	
stained glass	trainable lifeguard	mountain biking	
stationary making	trainee lifeguard	mountain boarding	
textiles	trainee teacher	open kayak	
Tye-Dye		outdoor adventure	
videography		paddle boarding	
wood working		paint ball	
		power boat driving	
		quad bikes	
		rack and field	
		rib driving	
		roller hockey	
		roller skating	
		rowing	
		shooting	

show jumping	
skate park	
ski boat driving	
soccer	
spinning	
swim coaching	
synchronized swimming	
tennis	
trapeze	
volleyball	
wakeboarding	
water skiing	
yoga instructor	
zip lines	
zumba instructor	

Table 3: Lemma candidate list for the field “Skills”

2.1.1.2.2 *Indeed* – a Job Search Engine as a Source for ESSCD's Notes

Indeed is a job search site, where instructor positions needed at camp are usually offered. The information on the website is used for the note sections of the ESSCD. In each entry of the Skills field there is a note *General tasks of* (name of the skill e.g. arts and crafts) *instructor*, which serves to prepare the users for the duties they will fulfill at the camp. Several job offers for arts and crafts instructors on the web page *Indeed* have been analyzed and essential duties and responsibilities have been summarized in a note. The descriptions of available work places on *Indeed* vary significantly: some provide a relatively detailed job specification and required qualifications, skills and experience, supplement functions as well as physical requirements whereas others just list the tasks and responsibilities.

2.1.2 Secondary Sources

Our secondary sources include audio materials, the Linguee English-Spanish Dictionary and additional grammar books. The headwords are recorded by English native speakers. The Linguee English-Spanish Dictionary assists with translation. Linguee is the largest multilingual online dictionary in the world, available in more than 200 language pairs[20]. Linguee is especially attractive for translators offering a number of parallel texts in the selected language pair. Linguee gives access to more than a billion of human translations. The translations which were not verified by professionals are marked with a warning sign.

2.1.3 Preliminary Conclusion

Corpora are the main source for the compilation of the ESSCD. It should be considered that crawling a corpus from the Internet usually involves a lot of noise, and words are frequently misspelled. In order to reduce the noise in the self-compiled corpus, external links were removed. Only the blog links were kept as a source of encyclopedic information. The English-Spanish Comparable Corpus tailored to the goals of our project, COCA, iWeb and the English Web Corpus 2015 form the text basis of the ESSCD, and act as the sources of collocations and corpus examples. The basis for grammatical information are grammar handbooks of each working language. Encyclopedic information is based on *Indeed* and the English-Spanish Comparable Corpus. Translation equivalents are taken from the Linguee English-Spanish Dictionary.

[20] Linguee.com: The World's Fastest Dictionary. Retrieved from https://www.linguee.com/press/EN/2015-02-09_PressRelease_EN.pdf

2.2 Dictionary Subject Matter

Wiegand (1998, p. 302) defines the subject matter of a dictionary as follows: "[d]er Wörterbuchgegenstand eines bestimmten Wörterbuches ist die Menge der in diesem Wörterbuch lexikographisch bearbeiteten Eigenschaftsausprägungen von wenigstens einer, höchstens aber von endlichen vielen sprachlichen Ausdrücken, die zu einem bestimmten Wörterbuchgegenstandsbereich gehören." The subject matter of the ESSCD is the summer camp application vocabulary. As has been mentioned before, summer camps are a part of the American and Canadian culture with a long tradition. Over 10 million children spend their summer in camps across the USA. Summer camp life is full of specific concepts, terms, activities etc. The ESSCD aims at describing them.

2.3 For whom is the ESSCD? User Profile

In order to satisfy user's needs, it is absolutely crucial to determine the profile of the consumer and the situations that trigger these needs: "[t]wo fundamental factors interact in the formation of [user] needs [...]. The first of these factors is the *characteristics of the concrete person* experiencing the information needs, and the second is the *social situation or context* where these needs occur" (Fuertes-Olivera & Tarp, 2014, pp. 48-49). Prior to making a dictionary, several competences of the target users are to be taken into account: mother tongue, which is"[t]he pivot in any user profile is **native language**[21]" (Bergenholtz & Tarp, 1995, p. 20), encyclopedic knowledge, native-language competence, foreign-language competence, native-language text production and reception, foreign-language text production and reception, translation from the foreign language into the native language and vice versa (Bergenholtz & Tarp, 1995, pp. 20-24). These aspects have been formulated in eight basic questions[22] suggested by the *function theory*[23] in order to build a profile of the target user (see Appendix 10). According to these questions, ESSCD's users are learners of English as a Foreign Language with Spanish as their mother tongue. User needs are linked to their knowledge. The intended users should have the minimum level[24] of English: A2 for support staff positions and C1 for camp counselor positions. This also determines the content of the ESSCD, i.e. it includes vocabulary which falls into the A2-C1 frame. The intended users should have basic skills in translating between the languages.

[21] Emphasis in original.

[22] The list of 8 questions was further extended to include 11 questions in Fuertes-Olivera and Tarp, 2014 (see Appendix 11) and is open for additional questions.

[23] The theory of lexicographic functions has been initiated and developed by researchers from the Center for Lexicography at the Aarhus School of Business since the 1990s.

[24] According to the Common European Framework of Reference for Languages.

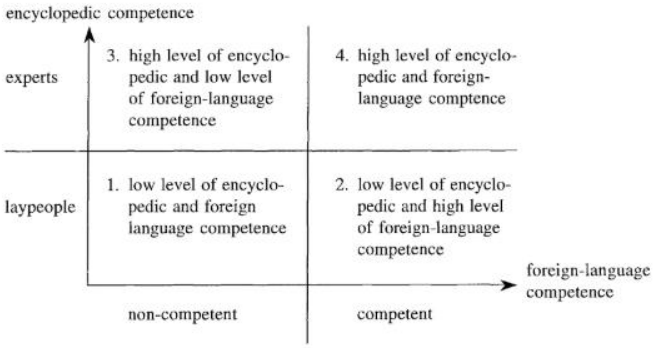


Figure 6: Levels of competence (Bergenholtz & Tarp, 1995, p. 21)

The participants are required to have at least a shallow level of the general cultural and encyclopedic knowledge. The intended users have to master the domain of summer camps at least at the beginner level:

Stage of application	Level of domain mastering
Users sign up for a program, i.e. get informed about the summer camp project, available positions, application process and receive additional information via email or read more on the web page of the sponsor/recruiter[25].	beginner
The candidate is hired and receives a set of educative videos about the summer camp life and his/her duties at the camp.	basic
The participant has lived the summer camp staff experience and wants to apply for a job for the next season. This type of applicant is called "returner".	semi-experts
Users have worked in a summer camp continually for over three seasons and more; applicants with working camp experience who are employed as recruiters and want to work during their vacations in the camp.	experts

Table 4: Levels of mastering the summer camp domain

Taking into account the encyclopedic knowledge of the users, the ESSCD is designed for laypeople and semi-experts. It is quite difficult to determine the mother tongue competence level of the intended users and the corresponding LSP in their mother and in a foreign language. In terms of *consultation-relevant* user characteristics (Fuertes-Olivera & Tarp, 2014, p. 50), the target users seem to lack experience in lexicographical consultations.

[25] Recruiters are self-employed workers or agencies that are in charge of promoting the program, inform students about the program and follow and support participants with the hiring process.

2.4 Usage Situations

This section describes concrete lexicographic situations that motivate users to consult the ESSCD. The motivation for a dictionary consultation, as a rule, arises from lexical gaps, the search for translation equivalents or spelling (Müller-Spitzer, 2014). As claimed by the modern theory of lexicographic functions, two main groups of use situations can be distinguished: *knowledge-orientated* and *communication-orientated*[26] (Bergenholtz & Tarp, 2003, p. 174). In knowledge-oriented situations, users seek additional information to widen their knowledge about the subject field in question, for example, encyclopedic information. In this type of situation, the communication occurs between the lexicographer and the user (the user seeks knowledge and the lexicographer provides it), whereas during communication-oriented situations, two or more persons are engaged in written or oral text production/text reception. In the latter case, the lexicographer takes the role of an indirect mediator and helps to solve communication difficulties, which arise during the encoding or decoding of texts. Five forms of communication are presented in the figure below. Communication issues may arise during the phases of production, reception and translation styled in italics.

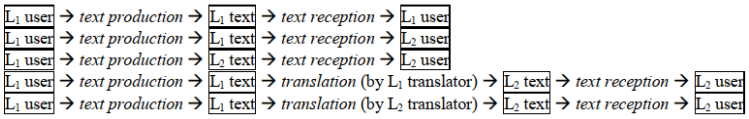


Figure 7: Communication model (Bergenholtz & Tarp, 2003, p. 174)

Keeping in mind the six types of communication-oriented use situation[27] suggested by Bergenholtz and Tarp (2003, p. 175), the ESSCD should be consulted under the following cicumstances:

- production of texts in English as a foreign language;
- reception of texts in English as a foreign language;
- translation of texts from English as a foreign language into Spanish as a native language.

However, the above-presented use situations, e.g. reception, production are broad concepts. Consequently, it is important to define the context of the dictionary use, i.e. the *extra-lexicographic* one (Tarp, 2012, p. 114), which will be described in the next section.

[26] The recent research of the Function Theory proposes four lexicographical situations: (1) communicative (e.g. text reception, production, translation etc.); (2) cognitive (users require information in order to save it in the memory as knowledge or to apply it for a particular task; (3) interpretive (users require information to interpret a special sign, symbol etc.); (4) operative ("a person needs information in order to perform an action of a physical, mental or linguistic nature" (Fuertes-Olivera & Tarp, 2014, p. 51).

[27] Six basic types of communication-oriented user situations (Bergenholtz & Tarp, 2003, p. 175) are listed in Appendix 12.

2.5 User Needs

Having defined the target users and the situations of use, we can proceed to the users' needs. Dictionaries are designed to meet the users' needs and to satisfy their preferences. For this reason, one of the leading tasks of the lexicographic team is to figure out the user necessities. User needs are connected to a specific user group and specific use situations (Bergenholtz & Tarp, 2003). Fuertes-Olivera and Tarp (2014, p. 48) point out that a distinction should be made between *information needs* and *lexicographically relevant needs*. The information needs "initially lead to a lexicographical consultation, but this consultation itself may give rise to a new type of need different from the former and related to the correct handling and use of a lexicographical tool, with the specific purpose of retrieving the information required to satisfy the original needs."

Nowadays there is a danger of information overload. We have too much data at our disposal and sometimes spend hours looking for a certain piece of information. The main reason for information overload in specialized dictionaries is the fact that the dictionaries regardless of their form try to satisfy many need types of varied user groups and propose all the possible solutions (Fuertes-Olivera & Tarp, 2014).

As Tono (2010, p. 3) points out "[n]o user has specific needs unless they are related to a specific type of situation". ESSCD is designed for certain users in particular non-lexicographic situations. Below is the description of the situations, when the user needs may arise and to which the ESSCD is linked. The user is willing to participate in a summer camp program, so he/she needs to fulfill the program requirements. One of the requirements is to fill in the application documents online. The application documents are available in English. Potential applicants may not understand the fields, so they would look for translations from English into Spanish. Furthermore, the intended user group may not be familiar with LSP terms in English, because they lack information about the summer camp subject field or some concepts existing only in the USA such as *Next of Kin's address*. Another problem that the target group might encounter is the lack of essential vocabulary to come up with exhaustive answers. The mentioned problems have been personally observed among participants when working for a placement agency. The user needs have been identified through personal communication with applicants and through guidance them in the hiring process.

Welker (2010) has reviewed 320 empirical studies on user research and concludes that the majority of studies involve limited number of participants. User research has been also criticized for the lack of usefulness and representatives. Tarp confirms the prospective of Sheatsley (1974) that the great number of surveys today is "a waste of time and money" (Tarp, 2009, p. 293). Apart from this, the

user tests carried out on the Danish Dictionary (Trap-Jensen, 2010, p. 1139) reveal that the actual users are usually "unable to analyze their needs". In this regard, the ESSCD will take a step forward[28] engaging potential users in questionnaires, protocols or tests with several samples of dictionary articles in order to better research user needs and to provide answers to the questions suggested by Fuertes-Olivera and Tarp (2014).

Not only should the dictionary itself be useful, but also the user should experience usability to obtain the maximum benefit from the consultation. The concepts of *usefulness* and *usability* were introduced by Laufer and Kimmel (1997). These scholars define usefulness as "the extent to which a dictionary is helpful in providing the necessary information to its user" and formulate the term "usability" as "the willingness on the part of the consumer to use the dictionary in question and his/her satisfaction from it" (Laufer & Kimmel, 1997, p. 362). The users want to consult a reference work when they find what they are looking for. It is often the case that the dictionary contains the desired data; however, the user is not able to find it. Therefore, the training of dictionary skills should not to be underestimated.

2.6 Functions and the Genuine Purpose of the ESSCD

The concepts of *lexicographic function*[29] and *genuine purpose* of the dictionary have been explained in the lexicographic discussion from two points of view: Wiegand's general theory of lexicography and the modern theory of lexicographic functions. The two theories agree that dictionaries are utility products and that lexicography is an independent discipline. Since the dictionary is believed to be a utility tool, human activities, situations and needs should be studied. The user situations have to be determined in order to specify the functions and the genuine purpose of the dictionary (Bergenholtz & Tarp, 2003). Wiegand (2001, p. 236) considers the genuine purpose of the dictionary to be a part of its function: "[d]er genuine Zweck ist ein genau bestimmter Teil einer Wörterbuchfunktion, und zwar gerade derjenige Teil, mit dem alle vorgesehenen Typen von Benutzungshandlungen und damit die vorgesehenen Typen von Suchfragen bestimmt werden, denn diese sind durch den jeweiligen Wörterbuchgegenstand determiniert", while the representatives of the Function Theory state that the "genuine purpose is made up by the totality of functions of a given dictionary and the subject field(s) that it covers" (Bergenholtz & Tarp, 2003, p. 176). As opposed to the theory of lexicographic functions, where dictionary functions are the main criterion for a dictionary typology, and dictionary functions shape the content of the dictionary, Wiegand (2001, p. 235) connects dictionary functions with the

[28] See Section 6.1 Limitations.

[29] The term *dictionary function* was introduced by Shcherba (1940).

dictionary type: "[...] mit der Festlegung einer Wörterbuchfunktion folgendes festgelegt wird bzw. gegeben ist. Gegeben ist der (meistens phänomenologisch bestimmte) Wörterbuchtyp, zu dem das Wörterbuch gehört, um dessen Funktion es geht [...]". The representatives of the Function Theory express their reflections regarding the cited quotation as follows: "the central difference between the two theories is that Wiegand includes a *dictionary type* in defining a dictionary usage situation and thereby in his definition of a dictionary function" (Tarp, 2008, pp. 93-94); "[...] he studies the dictionary "from within" (phenomenologically) and not from the point of view of the users and their needs in specific types of user situations" (Bergenholtz & Tarp, 2003, p. 193). The preparation of the ESSCD considers users and their needs above all, therefore the ESSCD is closer to the Function Theory in terms of its functions, user group and needs.

The Function Theory claims that the lexicographical function is "*the satisfaction of the specific types of punctual information need that may arise in a specific type of extra-lexicographical situations*" (Fuertes-Olivera & Tarp, 2014, p. 62); "is to provide assistance to a specific user group with specific characteristics in order to cover the complex of needs that arise in a specific type of user situation" (Bergenholtz & Tarp, 2003, p. 176). Fuertes-Olivera and Tarp (2014, p. 62) claim that functions are "foundation stones in any lexicographical concept" and "heart and soul of lexicography". In terms of its genuine purpose, the ESSCD is designed to assist summer camp applicants with submitting their application documents. Bergenholtz and Tarp (1995) suggest that bilingual specialized dictionaries should combine different functions, i.e. not just to contain translation equivalents. The communication-oriented functions of the ESSCD are:

- to assist the users in solving problems related to text reception in a foreign language;
- to assist the users in solving problems related to text production in a foreign language;
- to assist the users in solving problems related to translation of texts from a foreign language into the native language.

The knowledge-oriented function of the ESSCD is:

- to provide specialized information about the subject field to the users.

2.7 Preliminary Conclusion

Dictionaries are considered to be utility products. Therefore, it is of paramount importance to cater for the user needs. The ESSCD is compiled with a specific user profile in mind. The target group of the ESSCD are students in the age range between 18 and 29, who apply for work in American or Canadian summer camps. It is a specialized dictionary for laypersons and semi-experts. The users who have

already participated in the program are considered to be semi-experts, as they are familiar with application documents and the hiring process. The ESSCD is to be employed in three communication-oriented user situations[30]: (1) production of texts in a foreign language; (2) reception of text in a foreign language and (3) translation of texts from a foreign language into the mother tongue. The key functions of the ESSCD are to support text production and translation. The dictionary aims at providing accurate translation equivalents. Moreover, it should treat expressions such as rule collocations, which can be used by the participant to formulate convincing answers.

[30] The user situations are taken from the list of basic types of the communication-oriented user situations (Bergenholtz & Tarp, 2003, p. 175).

CHAPTER 3 ESSCD AS A HYBRID LEXICOGRAPHIC GENRE

A *hybrid dictionary* is defined as "[t]he combination of one or more types of REFERENCE WORK in a single product" (DoL, 1998, p. 69)[31] and is also called a "mixed, combined, compromise, blended, or mongrel" dictionary (Hartmann, 2005a, p. 195). Typical genres of hybrid dictionaries are *dictionary cum encyclopedia, dictionary cum grammar, dictionary cum thesaurus* etc. (see DoL, 1998, p. 69). Hartmann (2005a) in *Pure or hybrid? The development of mixed dictionary genres* analyses the above-mentioned genres and provides examples for further hybrids such as *dictionary cum film guide, dictionary cum picture book, dictionary cum concordance, dictionary cum lifestyle book* etc.

The dictionaries of today are getting more complex and combine various types. Hartmann (2005a, p. 194) names such hybrid dictionaries "multi-purpose reference works". The lexicographer highlights that nowadays many lexicographic systems incorporate spelling and thesaurus, thus the combination of various dictionary types is becoming a tendency. In the same vein, Nesi (2000, p. 839) states: "[a]s lexicographers and language educators explore new ways to present information about word meaning and use, the traditional distinctions between different categories of reference work are becoming increasingly blurred. Many recently published dictionaries are hybrids, merging features associated with more than one kind of wordbook, and taking on some of the duties of encyclopaedias, pedagogic grammars and teaching materials". Interestingly, Osselton (2000) employs the term "hybrid" for the *Oxford English Dictionary* compiled by Murray et al., the *Deutsches Wörterbuch* by Jacob and Wilhelm Grimm and others, for these dictionaries fulfill two functions: they describe the vocabulary of both the present and the past. Actually, some scholars (e.g. Shcherba, 1940) doubt the existence of conventional pure dictionaries, since many general language dictionaries also treat specialized vocabulary such as slang words and terms.

In order to satisfy specific user needs, i.e. to fulfill specific purposes, the ESSCD mixes five "pure" dictionary genres. The ESSCD is a bilingual specialized onomasiological online learners dictionary. The identified dictionary types are discussed in this chapter. The dictionary typology will receive thorough consideration.

3.1 Dictionary Typologies

Dictionary typologies organize these reference works according to particular criteria, e.g. the typology of Hausmann (1989) groups dictionaries focusing on their structure, Kühn's typology (1989) classifies dictionaries according to the usage functions, the typology of Engelberg and Lemnitzer (2009)

[31] Emphasis in original.

concentrates on the structure. Wiegand et al. (2010, p. 82f.) classify dictionaries according to the dictionary object (*Wörterbuchgegenstand*), dictionary form (*Wörterbuchform*), storage medium (*Speichermedium*) and publishing medium (*Publikationsmedium*). According to Hausmann's typology (see Figure 8), the ESSCD is a bilingual specialized dictionary.

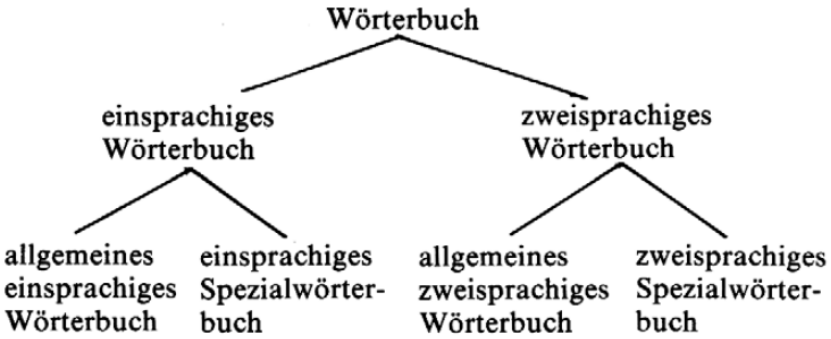


Figure 8: Dictionary typology proposed by Hausmann (1989, p. 973)

Taking into consideration Kühn's typology (see Appendix 5), the ESSCD should be used as a reference work (*Nachschlagewerk*) for text reception and production. In congruence with Engelberg and Lemnitzer[32] (2009), the ESSCD is a specialized dictionary with semantically restricted lemma selection, oriented at a particular user group (Wörterbuch mit semantisch beschränkter Lemmaauswahl; Benutzergruppenorientiertes Wörterbuch).

The above-mentioned typologies are medium-independent. In terms of electronic dictionaries, the typologies of Lehr (1996), Storrer and Freese (1996), Nesi (2000), Müller-Spitzer (2003) and de Schryver (2003) should be considered. Lehr (1996) calls for a distinction between electronic dictionaries that are based on paper dictionaries and those that are new developments, thus, this typology focuses on technical and (meta)lexicographic evaluation. Many electronic dictionaries are only digitized versions of hard copy dictionaries, converted without any updates (Nesi, 2000). According to the typology of Lehr (see Appendix 7), the ESSCD is a newly developed online dictionary with innovative design features. Nesi (2000, pp. 839-842) describes electronic dictionaries as retrieval systems and lists following dictionary types: the Internet dictionary, the glossary for on-line courseware, the learners dictionary on CD-ROM and the pocket electronic dictionary. Engelberg and Storrer (2016, p. 31) point out that online dictionaries combine information types in a new way, and describe Internet dictionaries as „Megawörterbücher" and „Wortschatzinformationssysteme", since modern dictionaries are merged

[32] For the typology proposed by Engelberg and Lemnitzer (2009), see Appendix 6.

with corpora, multimedia extensions and automatic language analysis tools. However, the typology of Internet dictionaries has not been fully elaborated until today.

Fuertes-Olivera and Tarp (2014, pp. 12-18) develop an unusual typology for specialized online dictionaries in the historical analogy with the invention of the Model T Ford and determine five categories: (1) Copycats; (2) Faster Horses; (3) Stray Bullets; (4) Model T Fords; (5) Rolls Royces. The first type are still printed dictionaries, but in electronic format, somewhat difficult to navigate. The given example is *Diccionario de la Lengua Española*. The Copycats group involves two subtypes "*old dictionaries* which are now merely used for research purposes and no longer as consultation tools, and *modern dictionaries* which are supposed to be used as reference works" (Fuertes-Olivera & Tarp, 2014, p. 13). Most of the existing online dictionaries belong to the *Faster Horses* category that "may be either electronic versions of existing paper dictionaries or completely "new" ones". In comparison to Copycats, Faster Horses enable quicker access and more advanced search options, although the dictionary articles are not dynamic. The indicated example is *Inter-Active Terminology for Europe* (IATE). The next type, the *Stray Bullets* either underestimates or overestimates the incorporation of new technologies. At one extreme, the dictionaries implement advanced techniques to a restricted extent. At the other extreme, dictionaries superfluously incorporate modern methods (the stated example is *EcoLexicon*). Consequently, technologies outweigh Lexicography, and its main objective to satisfy user needs loses its relevance. Fuertes-Olivera and Tarp highlight that the dictionary function is not meant to entertain the user, and the goal of Lexicography today is "to leave old habits behind and make full use of available technology in order to invent new advanced solutions to old problems" (Fuertes-Olivera & Tarp, 2014, p. 16). The last two types, *Model T Fords* and *Rolls Royces* stand for these new solutions, offering "*dynamic articles with dynamic data*, that is, articles that vary from consultation to consultation in terms of their lexicographical content" (Fuertes-Olivera & Tarp, 2014, p. 16). The essential purposes of *Model T Fords* are to meet the user needs in a specific situation or activity, so that the user can access data he/she is interested in on the Internet via hyperlinks. A proposed example for this type of dictionary is *Accounting Dictionaries*. *Rolls Royces* is a future type of online dictionaries designed to satisfy individual information needs of a concrete user in each consultation context. Fuertes-Olivera and Tarp compare Rolls Royces to driverless cars, where the users should only tell the vehicle the final destination. Hanks (2012) foresees that in the dictionaries of the future "contextualization and phraseology will come to take center stage. These dictionaries will be electronic products with hypertext structures and links, not printed books, nor the 'horseless carriages' that now pass for electronic dictionaries".

To sum up, dictionaries can be classified according to numerous criteria: from Kühn (1989) and Hausmann (1989) to Fuertes-Olivera and Tarp (2014). The great number of typologies have arisen due to the fact that a typology can be based on different criteria and parameters: "as the traditional image of the dictionary changes, new distinctions between dictionary types arise" (Nesi, 2000, p. 839). Some researchers even assume that "[d]ictionaries come in more varieties than can ever be classified in a simple taxonomy" (Béjoint, 1994, p. 37).

3.2 ESSCD as an Online Dictionary

The number of Internet users worldwide is steadily increasing: from 1,024 billion in 2005 to 3,57 billion in 2017 as recent statistics prove (see Appendix 8). This tendency has an impact on Lexicography as well. We are living the transition from printed to online dictionaries and searching an online dictionary is considered as a "revolutionary experience" (Nesi, 2000, p. 839). Nowadays, the users opt for ever faster ways of information retrieval, and online dictionaries provide this opportunity. *Electronic Lexicography* offers new possibilities: "new kinds of evidence, new modes of description, new ways of organizing evidence, new possibilities for exploiting database structure and hypertext links, and the need for new theoretical foundations" (Hanks, 2012). In addition to the term "Electronic Lexicography", Carr coins another term, *Cyberlexicography*, and defines it as "employing the Internet to compile or create a dictionary" (Carr, 1997, p. 209).

3.2.1 Onine vs. Print Dictionaries

> The best dictionary is probably the one rendering usable result in a short time.
>
> Bergenholtz and Gouws (2010, p. 114)

The opinions of scholars regarding the medium of dictionary production are split. On the one hand, online dictionaries were strongly criticized as unreliable sources in the middle of the 1990s: "[w]er glaubt, mit einem Internet-Anschluß auch eine Vielfalt von qualitativ hochwertigen Nachschlagewerken erworben zu haben, die die Anschaffung von Wörterbüchern und Enzyklopädien in gedruckter Form oder auf CD-ROM überflüssig macht, wird sich zunächst enttäuscht sehen. Das Internet mit seiner dezentralen Organisationsform und seinem schnellen Wachstum ist bislang kein Ort der Verbindlichkeit und Verläßlichkeit" (Storrer & Freese, 1996, p. 129). On the other hand, print dictionaries have obvious limitations, they "are likely to remain extremely conservative and command little or no serious investment and hence bring little or no serious innovation" (Hanks, 2012) and are "far from adequate as a medium for dictionaries" (Rundell, 2012, p. 16). In order to be an effective tool, both medium

forms should do the same: satisfy specific user needs within a short time and represent data as understandable as possible (Lew, 2012).

The significant increase in online dictionary consultations has been observed: "electronic learners' dictionaries already seem to be on the way to becoming a preferred alternative to the 'fat' dictionary in print" (Nesi, 1999, p. 65). This trend is growing thanks to innovative features of online dictionaries: portability, faster search, incorporation of multimedia elements (images, audio pronunciation, video clips etc.), mobile contents, immediate cross-references inside the dictionary and links to external sources, etc. New technologies speed up search queries and reduce the lookup time in online dictionaries as contrasted to leafing over the pages of their printed counterparts. These are the main reasons why many students favor the electronic medium (see the study of Taylor & Chan, 1994); the users appreciate "search speed and ease of use" (Dziemanko, 2012, p. 333). One of the latest developments that reduces the search time when reading a text in a foreign language is an electronic dictionary integrated into an application (e.g. Kindle). When the users come across an unfamiliar word, they can just click on it and a pop-up window with relevant lexicographical information will appear: "[i]t is already clear that the dictionary is moving from its current incarnation as autonomous 'product' to something more like a 'service', often embedded into other resources" (Rundell 2012, p. 29). Spelling and grammar checkers are seen as advanced lexicographical tools, because they detect mistakes and suggest corrections (Fuertes-Olivera & Tarp, 2014). On the other hand, distracting advertisements are present on the pages of some online dictionaries. Consequently, the consultation process might be slowed down.

The overwhelming majority of online dictionaries provide audio pronunciation which is considered to be a benefit, since not all users are able to interpret *International Phonetic Alphabet* (IPA) transcriptions. One more advantage of online dictionaries is the free of charge consultation, as opposed to their paper-based counterparts, which are usually quite costly. Furthermore, an online dictionary can be frequently updated, whereas users normally wait several years for a new edition of a hardcopy dictionary to be published. The internet dictionary allows users to interact with each other, to elaborate dictionary articles on their own or to improve the existing material[33] in collaboration with other users.

Online dictionaries can be accessed on any electronic device at any place and time, whereas their paper predecessors are usually bulky and difficult to carry around. The location of online dictionaries, i.e. URL, however, may change or disappear. Another option, offered by some online dictionaries, is the access to the most recent entries. Lew (2013) observes that paper dictionaries have *non-immediate*

[33] See Section 3.2.3 User Participation.

crossreferences, while electronic dictionaries mostly provide immediate crossreferences, so the users can "jump" from one article to another with a mouse click, since articles in an online dictionary are linked with each other. Besides, the majority of online dictionaries incorporate *error tolerant input*, which helps the users to look up words with wrong spelling and gives them the possibility to review the search history. Although online dictionaries introduce a wider range of access routes than printed ones do, the content quality is not assured: "[t]he range and convenience of such search routes in EDs are, of course, no guarantee of the quality of the information content" (Nesi, 2000, p. 840).

The size of a print dictionary leads to article density, therefore it is difficult for the user to retrieve information from condensed articles. In the print dictionaries era it was quite challenging to fit all the data on the paper pages. Therefore, many dictionaries were divided into volumes, which was not very practical for the user. Several volumes of a printed dictionary can be combined in one online dictionary: "one in which the electronic dictionary or dictionary site may encompass a single or a whole range of traditional dictionaries that can be adjusted in various ways to comply with the needs of particular user groups" (Trap-Jensen, 2010, p. 1133). In terms of space in online dictionaries, Lew (2013) distinguishes between storage, i.e. "capacity to hold the total content" and presentation space, i.e. "display of lexicographic information". Similarly, Trap-Jensen (2010, p. 1133) speaks about the *two-sided nature* of electronic dictionaries: (1) data that is saved in a database and (2) data represented on the screen. Regarding the presentation space, both dictionary forms (paper and electronic) have restrictions. Fuertes-Olivera and Tarp (2014) propose that the maximum amount of data that can be visualized in an online dictionary should allow the user to navigate without having to scroll down. Although online dictionaries have enough space to include images, tables or other additional data, users may experience *information overload*. This concept was introduced by the futurist Alvin Toffler in the book *Future Shock* (1970)[34]. Actually, users can experience information overload regardless of the medium. As mentioned earlier, the density in a print dictionary can also frustrate users. Reviewing hard-copy editions of the *Oxford Advanced Learner's Dictionary of Current English* (OALD4), Nesi (1999, pp. 55-56) states that "the more information the paper-based dictionary contains, the harder (and more time-consuming) it will become for learner users to find exactly what they need to know, without first having to negotiate a quantity of information that they do not need to know, or cannot process."

Now the question arises how to minimize the information overload and to identify "which particular e-lexicographic solutions work best (and for whom, and under what circumstances), so that future electronic dictionaries can be made more effective than their paper predecessors, and more effective

[34] Future Shock. In: Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Future_Shock

than the dictionaries available today" (Lew, 2012, p. 344). Fuertes-Olivera and Tarp (2014, p. 64) suggest possible solutions: "purely mono-functional; multi-functional allowing for mono-functional data access, mono-functional allowing for individualized data access or multi-functional allowing for individualized data access". Dictionaries should give the users the opportunity "to select, filter and present the specific data needed by the user" (Fuertes-Oliverera & Tarp, 2014, p. 93). These lexicographers take *Lexin*, online dictionaries with Norwegian as a source language designed for immigrants in Norway, as an existing example for designing individual articles. Lexin allows for the inclusion or exclusion of certain data items in the article[35], so that the problem of superfluous data may be solved, "electronic dictionaries [...] do not have the organisational and spatial constraints of hardcopy dictionaries, and can retrieve and combine information according to the specifications of the user" (Nesi, 2000, p. 839).

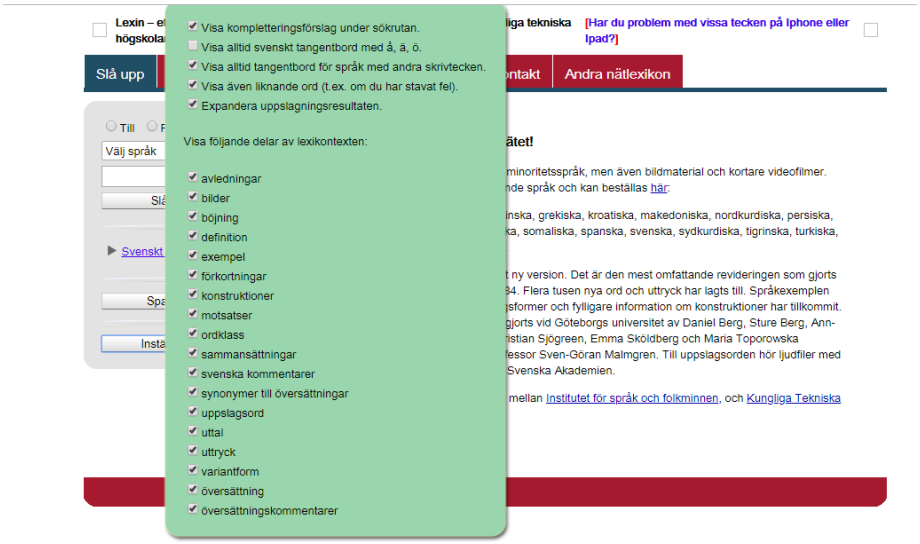


Figure 9: Pop-up window in *Lexin*

To avoid information overload, the ESSCD will provide the precise data required by specific users in specific usage situations. Given that the ESSCD is an onomasiological dictionary, the online production form is significantly important, because print dictionaries organize entries mainly in a linear manner, which is "inadequate as a means of grouping and regrouping words according to their semantic and pragmatic similarities" (Nesi, 2000, p. 839). The common characteristic of all reference works is that they are not meant to be read linearly, but searched for selective information retrieval (Engelberg, Müller-Spitzer & Schmidt, 2016). Furthermore, the target users are young people, who are accustomed to access the Internet in case of any information need. Rundell (2012, p. 72) refers to the users in the age range between 17 and 24 as "digital natives". From the lexicographer's point of view, online

[35] See Figure 9: Pop-up window in Lexin.

dictionaries are easier to edit, since a lexicographer does not need to review A-Z letters, but rather selects updates and additions. Thus, online dictionaries can be frequently updated. The conclusive argument in favor of creating an online dictionary for the ESSCD is the fact that the participants fill in the application documents online too. Therefore, it would be more convenient for them to access a new tab and start the dictionary search. As the first international study with 684 respondents on online dictionary use in 2010 reveals, the dictionary users are likely to consult dictionaries on laptops or computers rather than on the small screen devices (Koplenig & Müller-Spitzer, 2014). Having evaluated all the advantages and disadvantages, we have decided to compile the ESSCD in an online format.

3.2.2 User Research

There is a wide range of studies on comparing electronic and print dictionaries in terms of their effectiveness and usefulness. Töpel (2014) summarizes important individual studies on the use of e-dictionaries and concludes that users tend to consult electronic dictionaries more often, and find the required information faster than in print dictionaries. The users are generally satisfied with electronic dictionaries. It is worth mentioning that more surveys were carried out with CD-ROM dictionaries and PEDs[36] than with online dictionaries. The results of important individual studies on electronic dictionaries from 1993 until 2016 are visualized in Figures 10 and 11. Figure 10 shows that more than half of the informants (57,78%) experience difficulties using electronic dictionaries, i.e. they are not familiar with the innovative functions of e-dictionaries and, consequently, need training. This tendency can be observed during the whole period of time (see Figure 11).

[36] PEDs: Pocket Electronic Dictionaries; studies show that this type of electronic dictionaries is popular in Asia (see Boonmoh & Nesi, 2008; Chen, 2000; Boonmoh, 2012).

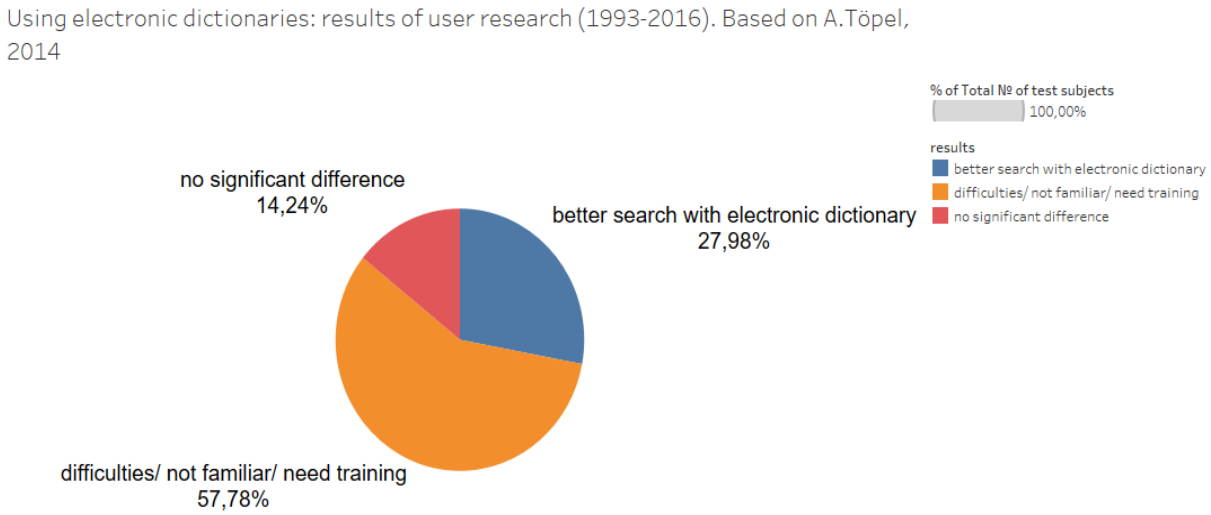


Figure 10: Results of important studies on electronic vs. printed dictionaries. Visualized in Tableau.

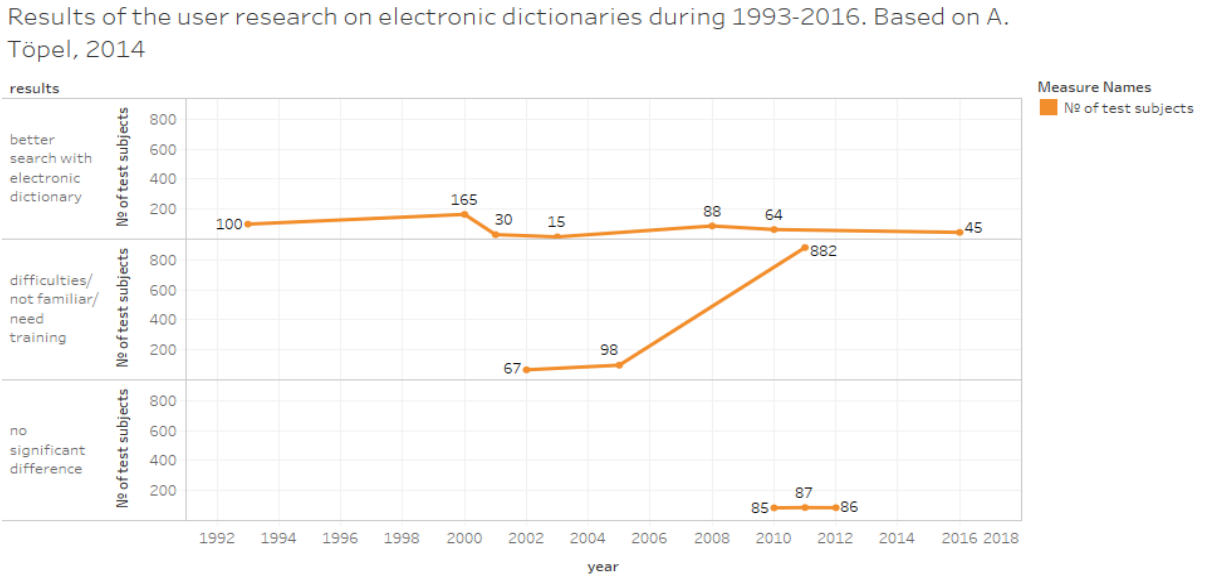


Figure 11: Results of important studies on electronic vs. printed dictionaries over time. Visualized in Tableau

It seems that users are not familiar with the novel features of e-dictionaries, unless they have been informed and trained on how to use innovative functions efficiently: "if a solution is unknown to the users, as is necessarily the case with any experimental feature we would like to test, their performance is likely to be negatively affected by the novelty of the feature. Depending on how steep a learning curve the new feature has, it may take more or less time and practice before users get more familiar with the innovation tested, and before the benefits, if any, get a chance to come to the surface" (Lew, 2011a, p. 11). The study by Taylor and Chan indicates that (1994) sometimes even English teachers have

questions regarding the use of electronic dictionaries. Even though the lexicographical data in online dictionaries implement innovative functions, the first two international studies on the use of online dictionaries prove that such novel features are not of greatest importance. The dictionary users still value conventional dictionary standards such as reliability of content, clarity[37], updated content, speed and accessibility (Müller-Spitzer & Koplenig, 2014). Regardless of age, occupation, language version, the results remain the same (Müller-Spitzer, 2016, p. 321f). Similarly Trap-Jensen (2010, p. 1142) reports "[...] it may be disappointing that the users do not seem to take advantage of all these wonderful possibilities." Reviewing the empirical surveys on innovative features of online dictionaries, Lew (2012, pp. 359-360) concludes that "available evidence invites optimism with respect to static pictures and audio recordings, but looks less optimistic when it comes to video and animation enhancements".

The focus of lexicographers on technical innovation has been criticized: "[..] it may be argued that the elements of customization implemented in electronic dictionaries so far result more from lexicographers' ideas about how users should use e-dictionaries (to the point that it might be called a 'lexicographer-oriented' lexicography) rather than from the insights into the way dictionaries are actually used" (Verlinde & Peeters, 2012, p. 151). Nesi (2000) states that the recent dictionary research concentrates rather on technical novelty than "the source, quality and appropriacy of the definitions" or other relevant aspects. An additional observation is that the majority of conducted user research is dedicated to the intra-lexicographical consultation[38] phase i.e. when the users are selecting information, accessing, verifying and retrieving data (Fuertes-Olivera & Tarp, 2014).

Summarizing the results on the survey "What makes a good online dictionary", carried out in 2010 and involving 1074 participants, Müller-Spitzer and Koplenig (2014, p. 184) conclude that "being a reliable resource, and a clearly presented and understandable tool, which is kept as up to date as possible" are the key features of a good online dictionary. Unlike paper dictionaries, modern Internet dictionaries integrate their user into dictionary development, offering a broader choice of user participation. This have led to a new form of dictionary making – *collaborative lexicography*, that will be discussed in the next section.

[37] Clarity is understood as "[t]he general structure of the website enables you to easily find the information you need" (Müller-Spitzer & Koplenig, 2014, p. 147).

[38] Fuertes-Olivera and Tarp (2014, pp. 87-90) determine three phases that make up the lexicographical process from the user's perspective: extra-lexicographical pre-consultation (the user experiences and becomes aware of the information need, as a result decides to start a dictionary consultation), intra-lexicographical consultation phase and extra-lexicographical post-consultation process (i.e. users apply retrieved information).

3.2.3 User Participation

Lexicographers can use the Internet not only for corpus searching or dictionary consulting, but also to reach out to their potential users. It is becoming more and more common that the Internet community contributes to the creation of new dictionaries and the improvement of already existing lexicographic resources. User participation is possible thanks to the following two main factors: the lexicographical process in online dictionaries is continuous, and online dictionaries are open systems. This phenomenon is called *bottom-up lexicography*. The term goes back to Carr (1997, p. 214) and constitutes the opposite to the top-down process where dictionaries are made from editors, through publishers, to readers. Lew (2014) proposes a new term for the modern user "*prosumer*" as a blend of the words "producer" and "consumer" for the users contribute to the dictionaries and profit from their content at the same time. Such user collaboration is significant for specialized lexicography in particular, including dialects, youth language or endangered languages. This section presents a review of interaction possibilities between the users and lexicographers.

Editors benefit largely from collaborative lexicography. They can gather first-hand information on user needs and preferences. The users can provide direct feedback in relation to the dictionary as a whole or to the single articles contained in it. In fact, the *Oxford English Dictionary* has a long tradition of collaborative editing. Its user participation practice reaches back into the 19th century when the *Philological Society of London* started working on the *New English Dictionary on Historical Principles*. The citizens of Great Britain, its colonies and North America were asked to collect and submit the examples for common word usage while reading: "[d]iese freiwilligen Helfer wurden gebeten, Bücher zu lesen, Belege auf Zetteln in einem bestimmten Format zu notieren und diese Zettel bei der *Philological Society* einzureichen" (Thier, 2014, p. 63). At the very beginning, the collection was unsystematic; further the public received the list of references that had already been processed and was asked to explore other text genres. Nowadays, users are encouraged to work on a specific domain e.g. current American texts, historical texts, scientific literature (see Their, 2014, pp. 63-64). Béjoint (1979) also encourages the involvement of users in lemma selection. The user contributions have increased the number of headwords in Collins, Longman and Cambridge dictionaries (Nesi, 2000).

Abel and Meyer (2016) determine three main types of user participation: *direct, indirect and accompanying* user participation (direkte, indirekte und begleitende Nutzerbeteiligung). Direct user participation refers to the possibility of the user creating, modifying and erasing dictionary articles. In the direct user participation, users can contribute to (Abel & Meyer, 2016, pp. 253-263):

- *open-collaborative dictionaries* (e. g. Wiktionary). Usually collaborative dictionaries have their predefined article structure, for instance, Wiktionary provides templates of Wiki Markup. In open-collaborative dictionaries, users are allowed to create, edit or delete articles.
- *collaborative-institutional dictionaries*. These dictionaries belong to a publisher e.g. Merriam-Webster Open Dictionary, Macmillan Open Dictionary. Unlike the previous category, the users can not directly edit or delete dictionary articles. Many user contributions in these dictionaries do not receive lexicographical attention.
- *semi-collaborative dictionaries* (e.g. OpenThesaurus, LEO). Lexicographers carefully check user contributions before including them into the dictionary.

The ESSCD, like any other reference work, strives to achieve that the users find what they are looking for. Hence, the ESSCD will implement indirect user participation[39], aimed at gathering feedback on existing or missing dictionary content, dictionary usage, or on the dictionary as a whole (Abel & Meyer, 2016, p. 263f). If users have questions about the functions or the content of the ESSCD, they are welcome to send an e-mail to the editorial board. Secondly, the user can fill in a blank form, where the request will be specified through the question and answer method. The processing of log files as implicit feedback is definitely too expensive for this project. Moreover, log files certainly do not output exact results as they do not distinguish between search engine access and users' access (see Verlinde & Binon, 2010).

The Internet dictionaries allow communication between users. The third type of user participation, namely accompanying user participation refers not only to the interaction between users and lexicographers (e.g. blogs, newsletters, language games), but also to the interaction between the users themselves. Fora, discussions within the user community, comments in social media are some examples of the interaction between users themselves. As a rule, the users ask for help with translation equivalents in bilingual dictionary fora. Interestingly, *Duden* offers linguistic consulting (*Sprachberatungsdienste*) on the phone, which is also considered a complementary participation (Abel & Meyer, 2016). The ESSCD will allow users to create their own user account, where they can mark their favorite words or multiword expressions with a star. This can provide data on the most read articles, motivating the lexicographers to improve the preferred articles. Another option will be to add lemmata to the "To Learn List" or marking them as "Already Learnt".

On the one hand, user participation might speed up the lexicographical process: publishing houses could save money, users become more familiar with the dictionary structure. On the other hand,

[39] See Section 6.2 Discussion on Future Directions.

dictionaries take a risk of including erroneous user contributions. As stated by Carr (1997, p. 214), "[t]he Internet is content-neutral: misinformation becomes concurrent with information". The result will depend on the quality of work users have done. Not everybody has a linguistic competence, consequently spelling and grammar mistakes or other errors often occur.

Regarding the quality of the user contribution, Abel and Meyer (2016) point out two quality issues: spam and vandalism on the one hand, and wrong, outdated or too complicated descriptions on the other. To hinder vandalism, Wiktionary, for example, provides detailed instructions to avoid copyright violation and encourages the users to report the cases of plagiarism. It also explains user rights and obligations. Besides, Wiktionary allows permanent access to the users who have written at least 200 articles. Hanks (2012) notes that the English Wiktionary contains outdated definitions, and suggests professional revision and corpora evidence in Wiktionary articles: "[i]n the English Wiktionary, the etymologies are taken from or based on those in older dictionaries; as are definitions, which are extremely old-fashioned and derivative" (Hanks, 2012, pp. 77-78). It turned out that these "stilted and archaic in wording" definitions were taken from the *Webster's Revised Unabridged Dictionary* (1913), which is available under public domain. Hanks also highlights the advantages of Wiktionary e.g. hypertexts to Wikimedia. Although Wiktionary is relatively new (the English language Wiktionary has its roots in December 2002), the number of user contributions to Wiktionary have increased rapidly[40]. Over 5 million users have written more than 30 million articles. Now the Wiktionary is available in 174 languages[41].

3.3 ESSCD as a Specialized Dictionary

Hartmann and James define specialized dictionaries as "[...] reference works devoted to a relatively restricted set of phenomena" and set them in opposition to general dictionaries: "[i]n contrast to the general dictionary which is aimed at covering the whole vocabulary for the 'general' user, special (or 'segmental') dictionaries concentrate either on more restricted information, such as idioms or names, or on the language of a particular subject field, such as the jargon of the drug scene or the technical terms of mechanical engineering" (DoL, 1998, p. 129). Fuertes-Olivera and Tarp (2014) made some observations[42] regarding this definition. In particular, the researchers note that Hartmann and James do not clearly distinguish between the terms *specialized dictionary* and *LSP dictionary*. Additional concepts

[40] A detailed study on the motivation of collaborative editing can be found in Abel and Meyer (2016).

[41] Wiktionary. In Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Wiktionary

[42] More on the discussion related to the definition of specialized dictionary see Fuertes-Olivera, P. A. & Tarp, S. (2014). *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlin: de Gruyter: 4-8.

introduced by Hartmann and James are *special dictionary* and *special-purpose dictionary* that contain only cross-references to the article *specialized dictionary* without additional explanation of these terms. Following the discussion on the above-cited definition, Fuertes-Olivera and Tarp (2014, pp. 7-8) use the term *specialized dictionary* to refer to "dictionaries, encyclopaedias, lexica, glossaries, vocabularies, and other information tools covering areas outside general cultural knowledge and the corresponding Language for General Purposes (LGP)". As can be seen from the definition, the range of specialized dictionaries is broad and embraces many reference works. As a consequence, the old distinction between dictionaries and encyclopedias is lost. The latter definition best fits the purposes of the current thesis since the lexicographical works that will be described and evaluated in the next section can be characterized as information tools.

An additional distinction within the specialized dictionaries distinguishes *culture-dependent* and *culture-independent* specialized dictionaries (Bergenholz & Tarp, 1995). It is difficult to say whether the ESSCD is culture-dependent or not. The subject of summer camps does not change radically within the country or language community. On the other hand, the users from the USA[43] would have a better understanding of the summer camp terms, because the camps are a part of the American culture. The ESSCD will provide a comparative description in case of cross-cultural differences.

3.4 ESSCD as an Onomasiological Dictionary

The core difference between *semasiological* and *onomasiological dictionaries* is that the former moves from the word to the definition, whereas the latter moves from the concept to the word. An onomasiological dictionary is defined as "[a] type of REFERENCE WORK which presents words or phrases as expressions of semantically linked CONCEPTS, which may be meanings, ideas, notions, word families and similar relationships" (DoL, p. 101)[44]. Sierra (2000, p. 224) summarizes the alternative terms for onomasiological dictionary: "ideological dictionary [Shcherba 1940], semantic dictionary [Malkiel 1975], conceptual dictionary [Rey 1977], speaker-oriented lexicon [Mallinson 1979], thematic wordbook [McArthur 1986], nomenclator [Riggs 1989]". According to Hartmann (2005a, p. 194), thematic order is based on "a topical system of knowledge". As explained earlier, the ESSCD is based on ten topics of the summer camp application. Especially for bilingual dictionaries (as is the case of the ESSCD), the onomasiological arrangement is adequate, because it restricts the list of possible translation equivalents and, as a consequence, reduces potential ambiguity. Instead of offering a string

[43] American users do not belong to the target group of the ESSCD.

[44] Emphasis in original.

of one-word translation equivalents or near-synonyms, the ESSCD usually provides only one equivalent. The alphabetical arrangement of dictionaries has been considered inadequate because it does not show the relationship between words and also complicates the search for fixed expressions (see Nesi, 2000). Since the ESSCD is an online dictionary, a combination of alphabetical and thematic access is possible. In order to see how online onomasiological dictionaries implement the combination of index based search with other search options, two online onomasiological dictionaries will be analyzed in the next section.

3.4.1 Analysis of Online Onomasiological Dictionaries

Computer and Internet technologies are advancing rapidly, and many online dictionaries have been launched during the last years. Onomasiological dictionaries are among them. Initially, we aimed to examine bilingual thematic online dictionaries. However, those are usually designed as picture dictionaries. Given that existing online projects working on bilingual thematic dictionaries are still in progress (e.g. Italian-English Dictionary of Collocations[45] and German-Spanish DICONALE of verbal lexemes[46]), this section discusses monolingual onomasiological online dictionaries, namely *BerufeNet* and *Computer Glossary*. The main criteria for the selection of these dictionaries were that they are created and managed in digital form, and not digitized versions of print dictionaries, and incorporate innovative search features, including onomasiological search. Interestingly, the dictionaries not only define themselves as dictionaries, but also as tools. Computer Glossary reports that it is "[t]he Tech dictionary and IT Encyclopedia; reference and self-education tool about information technology"[47]. BerufeNet is an information portal of the German Federal Employment Agency (*Bundesagentur für Arbeit*) about professions in Germany[48]. These dictionaries once again reveal the above-cited quote of Rundell that today's dictionaries are rather incorporated into other resources than are independent reference sources. BerufeNet is embedded into the web page of the Federal Employment Agency and Computer Glossary into TechTarget[49]. Both dictionaries are specialized dictionaries targeted primarily at experts of particular domains. The evaluation of both dictionaries will be based on parameters proposed by Kemmer[50] (2010, p. 30). Special attention will be devoted to dictionary search capabilities.

[45] The project is described in Berti & Pinnavaia (2012, 2014).

[46] For more detail see Sánchez Hernández (2013); Meliss & Sánchez Hernández (2015).

[47] About Us. In Computer Glossary. Retrieved from https://whatis.techtarget.com/about

[48] Begriffserklärung: BERUFENET. In Bundesagentur für Arbeit. Retrieved from https://www.arbeitsagentur.de/lexikon/berufenet

[49] TechTarget is an editorial and marketing company. Available at https://www.techtarget.com/

[50] See Appendix 9: Parameters for the description and evaluation of online dictionaries (Kemmer, 2010, p. 30).

3.4.1.1 *BerufeNet* – an Information Portal on Professions in Germany

BerufeNet is an expert database of the Federal Employment Agency for training and job descriptions; a free online information portal for all occupations known in Germany, principally used in career orientation and job placement (Dengler, Mattes & Paulus, 2014, pp. 10-11). BerufeNet integrates user guidelines in an innovative and efficient way. Every time the user selects a search option, a brief description of its function is provided. Furthermore, the dictionary incorporates a large number of question mark icons, which can clarify user doubts or explain functions of a particular search option. This concept of user guidelines becomes quite useful since "[t]he general assumption is that no-one bothers to read the front matter of dictionaries" (Kirkpatrick, 1989, p. 754). Once users encounter difficulties, they can resolve them directly, and there is no need to access the Frequently Asked Questions (FAQ), for example. Every professional field and all the sub-fields incorporate question mark symbols, which appear as a pop-up window with the explanation of the field/sub-field and its key points. People with disabilities can become familiar with the project and the navigation throughout the dictionary thanks to a special feature of Federal Employment Agency web site "*Leichte Sprache*"[51] that explains the dictionary in an "easy or simplified language" according to the rules of *Inclusion Europe*[52]. The metainformation about the dictionary is provided in the outer texts.

The outer texts in BerufeNet are situated on the right-handside of the screen and consist of three windows:

- New in BERUFENET (*Neues in BERUFENET*) informs about new versions, functions and content.
- Recent news (*Aktuelles*).
- Changes in the professional world (*Änderungen Berufswelt*) that is divided into three parts. *Neuordnungen Berufe* informs about reorganizations of occupations during the last years. Here the user can select the period he/she is interested in to track the changes. *Aufgehobene Berufe* lists extinct professions. *Änderungen Berufsbezeichnungen* arranges renamed job titles since 1996 alphabetically, and presents them in two columns: former job title and new job title.

When the user starts browsing the dictionary, a new window "*Zulezt Gesehen*" (last viewed) appears. Quite surprisingly, the dictionary does not report on its sources as a whole, each article has its own detailed list of information sources instead. The main source for lemma candidate list was the classification of occupations (*Klassifikation der Berufe 2010*).

[51] BerufeNet in Leichter Sprache. Retrieved from https://www.arbeitsagentur.de/leichte-sprache/berufenet-in-leichter-sprache

[52] Inclusion Europe. Available at https://www.inclusion-europe.eu/

BerufeNet contains nearly all job titles and their fields of specialization as well as courses of study, vocational and further training. The database incorporates more than 300 000 job titles, including male, female and neutral job titles, synonyms as well as English and French job titles (Janser, 2018).

As a rule, the dictionary comprises single word lemmata. Depending on the search option (professional field, occupation field, subject etc.), the lists of fields and sub-fields of the dictionary differ. The professional field contains all the headwords of the dictionary. Due to space constrains, only the general overview of the professional fields will be listed:

- Landwirtschaft, Natur, Umwelt
- Produktion, Fertigung
- Bau, Architektur, Vermessung
- Metall, Maschinenbau
- Elektro
- IT; Computer
- Naturwissenschaften
- Technik, Technologiefelder
- Wirtschaft, Verwaltung
- Verkehr, Logistik
- Dienstleistung
- Gesundheit
- Soziales, Pädagogik
- Gesellschafts-Geisteswissenschaften
- Kunst, Kultur, Gestaltung
- Medien

The microstructure of the BerufeNet dictionary articles differs from the one of conventional language dictionaries. It does not provide any grammar labels, collocations or examples. BerufeNet categorizes articles according to the six occupation groups: training occupations (*Ausbildungsberufe*), professions that require university education (*Hochschulberufe*), careers in public sector (*Beamtenlaufbahnen*), further training occupations (*Weiterbildungsberufe*), fields of study (*Studienfächer*) and military careers (*Militärlaufbahnen*). The article microstructure consists of lemma sign and five tabs. The lemma sign combines the masculine form and feminine suffix or *-frau* separated with slash, for example *Sänger/in, Schauspieler/in, Alleinsteuermann/-frau.* Some feminine forms are written in their entirety, e.g. in case of vowel change as in *Koch/Köchin, Anwalt/Anwältin* or other cases where adding a suffix is not enough to build a feminine form, i.e. further modifications will be needed, for example to delete "e" like in *Pädagoge/Pädagogin*. The full forms are listed to avoid confusion.

The graphical user interface of articles consists of five tabs:

- ✓ Brief descriptions (*Kurzbeschreibungen*). The brief description reports about the job activity at a glance and lists typical industries in which employment can be found.
- ✓ Admission requirements[53] (*Zugangsvoraussetzungen*). The admission requirements are subdivided with a drop-down menu into technical and personal requirements. The technical

[53] For discussion on admission requirements see Jansen (2018).

requirements inform about the necessary degree, recognition of foreign qualifications, other school or academic knowledge etc. To the personal requirements belong health, interests, work behavior, social behavior or other abilities.

- ✓ Education (*Ausbildung*). The education tab informs about the possibilities for further training, its duration, conditions and financial support as well as similar further education options.
- ✓ Professional activity (*Tätigkeit*). The professional activity tab employs a drop-down menu with the items Job Description and Conditions, sometimes also media (e.g. usually ten-minute videoclips called "Berufsfilm"). Job Description advises about job content, prospective financial benefits, job titles and competences that may be relevant to the practice of a particular profession. Conditions describe work conditions (descriptions of the work place and situations), work equipment and a detailed list of industries where jobs can be found.
- ✓ Prospects (*Perspektiven*). The prospects tab has a drop-down menu with further information on training, alternatives, job market and statistical data (see Figure 12).

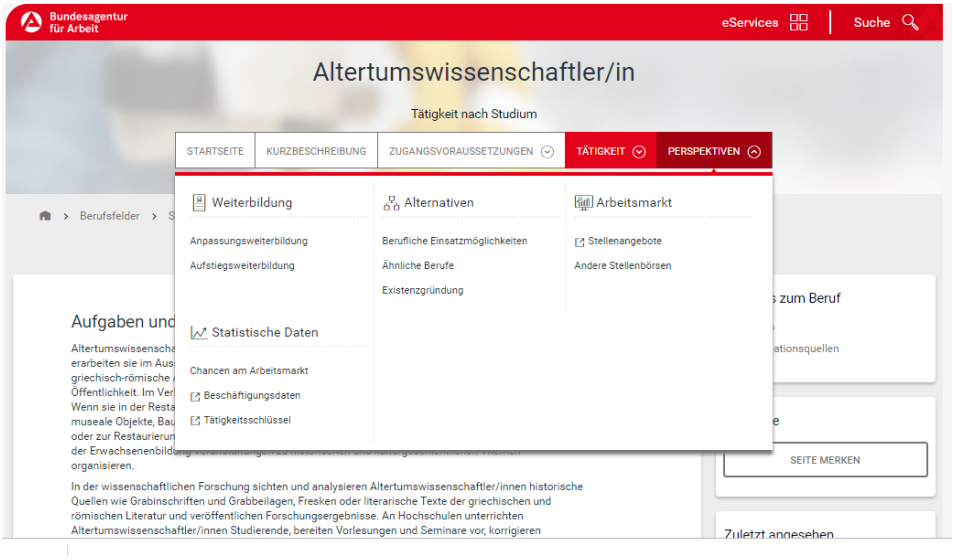


Figure 12: Drop-down menu of the Tab "Perspectives"

It goes without saying that the articles of BerufeNet are rich in encyclopedic data. The overwhelming majority of articles incorporates images that illustrate work in action. When a technical term which is likely to cause comprehension problems occurs in the article, it is highlighted in red. When the user clicks on the highlighted word, a pop-up window with an explanation appears (e.g. see the article Agrarbiologe/-biologin).

Although the microstructure of articles is quite complex with many tabs and drop down menus, BerufeNet is easy to navigate, because it shows the user the respective search path, e.g. Tätigkeitsfelder>Suchergebnisse Tätigkeitsfelder > Kurzbeschreibung > Arbeitsbedingungen. BerufeNet

articles can be accessed through the seven main search techniques and an advanced search, which implies filters (see Figure 13).

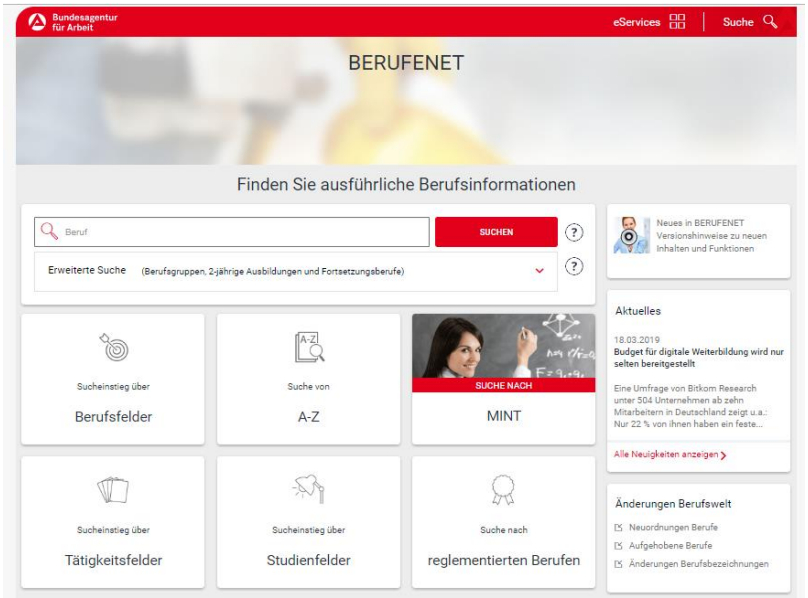


Figure 13: Start page of *BerufeNet*

Firstly, users can type a job title in the input field. Having typed three characters, the system proposes possible options. It seems that the job titles are connected to an occupation: when the user types a general profession in the search field, for example, Friseur, the query obtains 8 results (among them *Friseur*), which can be sorted alphabetically or by professional group, i.e. Berufskundliche Gruppe (see Figure 14). The next search function is an alphabetical one. The users should select the letter to see all professions and subjects beginning with that letter. The range of occupational activities, employment opportunities and university subjects is very large. Therefore, BerufeNet supports the search via professional (*Berufsfelder*) and occupation fields (*Tätigkeitsfelder*). As mentioned above, the former search option organizes all lemmata (professions, vocational trainings and fields of study) by topic (e.g. health, media, services). In the latter search option, users can find concrete occupations and job opportunities that fit their professional plans and expectations after the graduation or training completion.

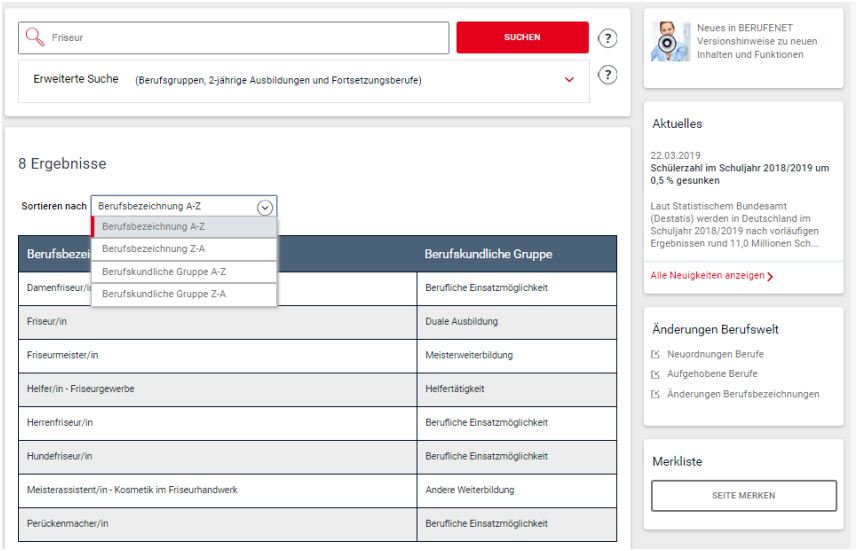


Figure 14: The results of the search query *Friseur*

The search may be also conducted according to the subjects, which are assigned to one or more fields of study. This option helps the users in the step-by-step search for a specific field of study. In particular, dictionary articles on courses of study inform users about the enrollment requirements, course contents, financial aspect, length of the study and list universities that offer a chosen field of study and employment possibilities after graduation.

The MINT-Search[54] option outputs professions and subjects, for which mathematics knowledge is of great importance. The last search option, *Suche nach reglementierten Berufen*, is performed according to regulated professions such as medical, legal, teaching professions, occupations in the public service. This means that these types of jobs depend on governmental acknowledgment for professional qualifications.

BerufeNet is being updated continuously. As a rule, the updates are based on requests of federal employment agencies and on official sources, for example, training regulations. The feedback from employers, employees and other public institutions also supports development of BerufeNet (Janser, 2018). If questions concerning the functions or content of BerufeNet arise, the users are encouraged to send an e-mail or call the editors directly. The current version number can be found on every page in the footer.

There is also a possibility of adding dictionary articles to a watch list (*Merkliste*), which appears as a separate window on the left-side of the screen and allows users to access already marked articles immediately. Moreover, BerufeNet incorporates a wide range of internal links to job trends, data

[54] MINT stands for Mathematik, Informatik, Naturwissenschaften and Technik (mathematics, informatics, natural sciences and technology).

sources, legal regulations (as every single state in Germany can have its own regulations) etc. and external links to related web pages, documents, brochures etc.

The dictionary is primarily designed for the workers of employment agencies: "[t]he purpose of this database is two-fold: it is used by vocational counselors and job placement officers at local employment agencies for career guidance and job placement, but it also serves the general public as a free database for career orientation" (Janser, 2018, p. 18). Looking at the content coverage of BerufeNet, it seems to be especially useful for high school graduates. BerufeNet provides detailed information about possible subjects to study, entry requirements and career paths. University or vocational training graduates can also benefit from BerufeNet by becoming familiar with an exhaustive list of potential work places and requirements. This information system provides young people with some idea of what to expect from their future jobs and what the working environment looks like. Furthermore, people with work experience are likely to make use of the dictionary as they might be interested in further education, job market, alternative professions. The BerufeNet can be compared to the American Dictionary of Occupational Titles (DOT) by the US Department of Labour. One of the significant drawbacks of the DOT[55] is that it is not kept up to date, even though its web page states that it was revised in April 2011. Like all paper dictionaries, DOT provides a Table of Contents, which is considered to be outdated. When users want to access articles thematically, they have to go through a long search path to access them. The structure of articles in DOT is not clear. The dictionary interface is not user-friendly, since all the contents of the dictionary articles are located mainly on the left-side of the web page

To sum up, BerufeNet is an innovative dictionary with a rich content. It provides detailed information on the current job requirements, prerequisites, tasks to fulfil in a particular occupational field, working equipment, the working conditions, as well as numerous professional alternatives. The dictionary offers a wide range of lookup options and a user-friendly interface. Although the dictionary contains a large amount of data, this data is clearly structured.

3.4.1.2 Computer Glossary

Computer Glossary is "the IT encyclopedia and research library in TechTarget's network"[56]. It consists of two parts "Browse Definitions by Topic" and "Quick Study Resources". The current description focuses on the first part of the dictionary.

[55] Also called Occupational Information Network (O*NET).

[56] Business Partners. In Computer Glossary. Retrieved from https://whatis.techtarget.com/about/partners

The dictionary provides users with the definitions of more than 10 000 terms and around 1 000 fast references, quizzes and cheat sheets. Computer Glossary is daily updated. The terms in the glossary can be accessed via the following "topics":

- Application Development
- Business Software
- Computer Science
- Consumer Tech
- Data Center
- IT Management
- Networking
- Security
- Storage and Data Management

Each topic is subdivided into various sections.

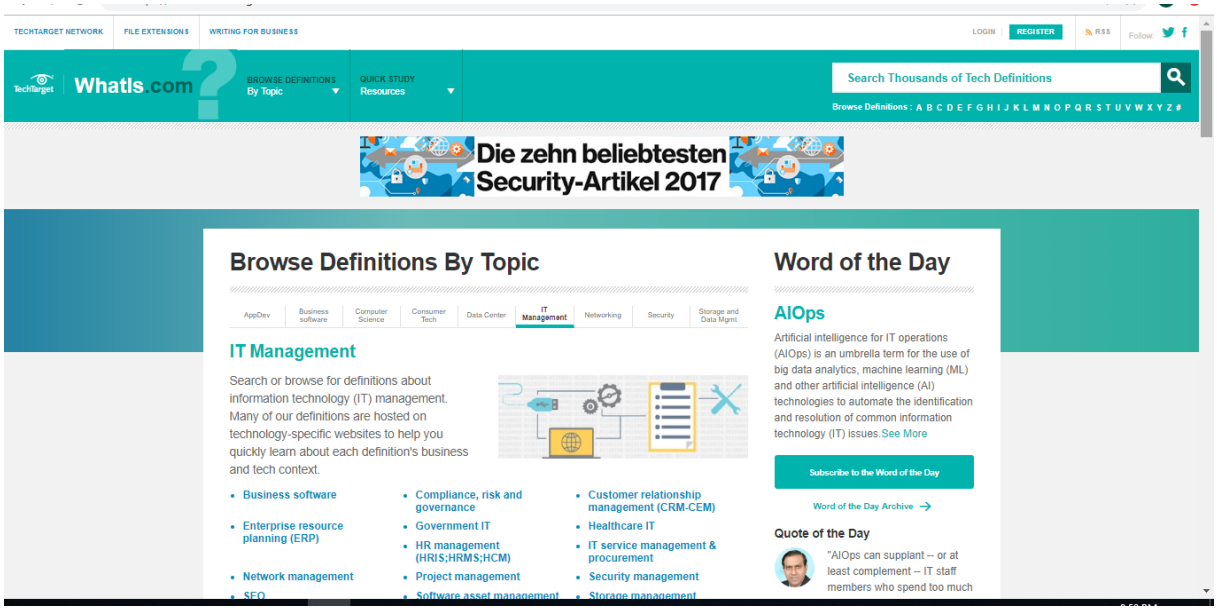


Figure 15: *Computer Glossary* start page. Subfields highlighted in blue.

The target group of the Computer Glossary is information technology and business professionals. Various abbreviations such as AppDev instead of Application Development, Data Mgmt instead of Data Management prove that the dictionary is aimed at experts. The genuine purpose of Computer Glossary is to assist IT specialists and TechTarget customers with comprehension issues, which are usually technology companies such as Dell, Intel, Cisco, Microsoft. In order to provide an excellent Business-to-Business service (also called B2B), both parties should understand each other. Since IT and technology

professionals use highly specialized vocabulary, the dictionary definitions should explain technical terms and business concepts clearly and in a compact way.

The outer texts of Computer Glossary are situated on the right-hand side and on the bottom of the web site. They include:

- Word of the Day. Users can subscribe to the Word of the Day and see the Word of the Day archive.
- Quote of the Day
- Meet the Editor
- Recently Published Definitions
- Newest and Updated terms
- Quiz Yourself
- Trending Topics
- Buzzword Alert. This feature shows a word which is fashionable at the moment of lookup.
- View Related Content. At the end of the dictionary articles there is a suggestion of further reading.
- Dig Deeper on e.g. telecommunication networking
- Related Terms

Users are encouraged to participate in dictionary editing in many ways. Firstly, in the *Meet the Editor* window, the users are asked to share their feedback, to suggest a new term or an update. The photo of the editor and the title "Meet the Editor" convey an impression that lexicographers are closer to the users, making it easier to start collaboration. Moreover, even the editor may initiate a discussion. Below the dictionary article there is usually a question from the editor and an invitation to join the discussion, followed by a list of related terms and a list of sources which can be read by users for additional information. Another option of user contribution is the forum *Ask Your Peers a Question*. On the bottom of the page the users are encouraged to comment dictionary articles with imperatives "Start the conversation", "Share your comment". Users can receive notifications when other members comment the same article.

The dictionary incorporates internal and external links to TechTarget's other IT-specific Web sites. Computer Glossary does not provide a user manual, images or the incorporation of sound files. The advertisement of books or services provided by TechTarget appears on the dictionary page. However, these advertisements are not as bright and annoying as the regular Web advertisement. Like BerufeNet, Computer Glossary provides the user with the search path followed to obtain a specific

result: from the start page to a particular article. Both dictionaries, BerufeNet and Computer Glossary, are easy to navigate; the users can always maintain the overview of their search actions. However, BerufeNet incorporates a broader variety of visual means such as images, font type and size that, in our opinion, facilitate information retrieval and help to avoid the "lost in hyperspace" effect (Kemmer, 2000, p. 16). It is also worth mentioning that the design and content of the analyzed dictionaries is adaptive to mobile device consultation.

Computer Glossary offers several search modi. Users can browse definitions by index. Compared to BerufeNet, where alphabetical search results are presented in a two column table (professional title and professional group correspondingly) and the headwords are situated on several pages, Computer Glossary splits the results of one letter into "subgroups" (e.g. A – ACC, ACC – ACT, ACT – ADV etc.) on the left-hand side as a scroll down menu. Phrases in Computer Glossary are presented as separate lemmata. The user can also search terms typing them in the search field. However, the automatic term completion[57] from an index is not available as is the case of BerufeNet. Notwithstanding, Computer Glossary incorporates a search technique with a combination of two or more words and allows the results to be sorted by relevance or date. None of the examined dictionaries implements spelling-tolerant search.

Each dictionary article in the Computer Glossary consists of lemma, definition and the date when the article was created or updated for the last time. Apart from defintions, many dictionary articles include encyclopedic information (e.g. common features, benefits, comparison with similar terms). Computer Glossary also provides some country-specific information, for example: "A-Law is the type of PCM used in most of the world. The other type, mu-Law, is used in the United States and Japan." The dictionary indicates who posted a certain article and who contributed to it. The following observations can be made regarding the lemmata. Firstly, editors seem to be inconsistent with abbreviations which are a part of lemma. Sometimes the acronym is placed before the corresponding full form in parenthesis, for instance *AUI* (*attachment unit interface*), sometimes the other way around: the full form precedes the acronym in parenthesis e.g. *Access Network Query Protocol (ANQP).* Furthermore, a headword is sometimes composed of two synonyms, for example, *cable TV or CATV (community antenna television),* even though the *CATV* already exists as a separate article. The internal link from cable TV to CATV is implemented in the definition. The alternative terms for the designation of the same concept are separated with "or" as in the example *cable TV or CATV* and parenthesis e.g. cloud telephony

[57] Also called *type-ahead search*, *search-as-you-type*, *incremental search*, *inline search*, or *instant search* (Lew, 2013, p. 24).

(cloud calling). In order to achieve consistency in all articles, editors should follow standardization principles.

3.5 ESSCD as a Bilingual Dictionary

Bilingual dictionaries have a long tradition. In fact, a bilingual dictionary is older than a monolingual one. For example, the first bilingual dictionary for the language pair English-Latin appeared before 1450, English-French in 1570, whereas the first monolingual English dictionary was published in 1604 (DoL, 1998). A bilingual dictionary is defined in contrast to a monolingual one as follows: "[a] type of DICTIONARY which relates the vocabularies of two languages together by means of translation EQUIVALENTS, in contrast to the MONOLINGUAL DICTIONARY, in which explanations are provided in one language"[58] (DoL, 1998, p. 14).

A bilingual dictionary is sometimes called a translation dictionary, however, the second term rather refers to the conventional type whose main function is to provide translation equivalents (Stark, 2011). Regarding the access to translation equivalents, bidirectional and monodirectional dictionaries can be distinguished. The translation equivalents in bidirectional dictionaries can be reached from L2 to L1 and vice versa (DoL, 1998). The ESSCD is designed as a monodirectional dictionary. The user can look up words from English to Spanish. The fundamental reason for this is that the applicants receive the application documents in English. Bidirectionality may be suitable once the dictionary is further developed and goes outside of the frame of the application documents, i.e. has a high number of lemmata. On the other hand, Bergenholtz and Tarp (1995) suggest the compilation of bilingual specialized dictionaries as monodirectional, because some concepts do not always find their equivalents in other languages.

Compared to monolingual dictionaries, the bilingual dictionaries have two absolute advantages. Firstly, they provide translation equivalents. Secondly, they focus on the particular language pair, thus considering the special features of these languages, e.g. false friends. On the other hand, the syntactic and morphological characteristics are not explicit enough. Definitions and examples are not as comprehensive as in the monolingual dictionaries (Lemnitzer & Engelberg, 2009). Bogaards (1996, p. 300) reports that monolingual dictionaries are more useful for reception than for production: "If you do not already know the L2 item you want to investigate, how will you find it in an MLD?"

Although it is generally accepted that language learners would benefit more from monolingual dictionaries, they usually opt for bilingual ones. One of the broadest studies in learners dictionary use

[58] Emphasis in original.

(Atkins & Knowles, 1990) with more than 1000 participants from seven countries has proven that the language learners mostly (75%) use bilingual dictionaries. However, this does not imply higher efficiency. The survey actually demonstrated that the use of monolingual dictionaries for task fulfillment leads to better results. In contrast, Lew's survey (2016) revealed that active bilingual learners dictionaries significantly improve the quality of writing (for about 33%). The surveys of Laufer and Melamed (1994) have proven that efficient dictionary use strongly depends on the user competence and dictionary search abilities. Competent users scored better with monolingual dictionaries whereas unskilled users with bilingual ones.

Stark (2011) considers three main advantages of the bilingual thematic dictionaries over the monolingual ones. First of all, it is much easier for the users to follow information in their mother tongue. In addition, bilingual dictionaries can resolve language-specific difficulties such as false friends. Last but not least, it is sometimes difficult to provide satisfactory definitions in the same language. Béjoint and Moulin (1987) point out that the bilingual dictionaries are suitable for a brief consultation, whereas the monolingual ones represent the lexical system of a foreign language. As mentioned before, the language learners, especially beginners, avoid consulting monolingual dictionaries and find them too difficult, albeit the foreign language teachers always encourage their students to use the latter. Moreover, think-aloud protocols from Wingate's survey (1999) suggest that the participants using the monolingual printed dictionaries sometimes forget which word they were looking for. Lexicographers differentiate between active and passive bilingual dictionaries that will be introduced in the next section.

3.6 ESSCD as an Active vs. Passive Dictionary

The idea to classify dictionaries as active and passive goes back to Shcherba (1940). Engelberg and Lemnitzer (2009, p. 126) summarize the difference[59] between active and passive dictionaries as follows: "[d]as aktive Wörterbuch erfordert eine umfangreiche Mikrostruktur, das passive eine umfangreiche Makrostruktur". In the lexicographic reality we can hardly find examples which would stand as prototypes for active or passive dictionary implementation model. Hence, it is more adequate to use terms active or passive dictionary functions (Mugdan, 1992). As Frankenberg-Garcia (2015) highlights, learner's dictionaries do not separate examples which support comprehension and those which facilitate production.

The survey on usefulness of bilingual thematic dictionaries (Stark, 2011) shows that 45% of English as a foreign language learners use this dictionary type for writing. The ESSCD is primarily designed as a

[59] For more on the differences between active vs. passive dictionaries see Engelberg and Lemnitzer (2009, pp. 125-133).

production dictionary. As explained above, the ESSCD combines both types, as its main purposes are to assist the learners with translation (decoding) and text production in the English language (encoding). The passive and active functions of the ESSCD have been described in Section 2.6 Functions and the Genuine Purpose of the ESSCD. It is commonly agreed that the text production process is more challenging for language learners than the comprehension one (Rundell, 1999). Lew (2013) stresses that very few dictionaries are designed as production resources and explains the reason for this. Firstly, this dictionary type should have richer content and, as a consequence, requires more lexicographical effort. Secondly, paper-based dictionaries do not have enough space for "redundant" content. Engelberg and Lemnitzer (2009, pp. 120-121) distinguish four reasons for dictionary consultation for text production:

- *Lexeme usage* e.g. valency, collocations, inflexion, connotations. The best option to solve this problem is a learner's dictionary.
- *Lexeme finding* (passive vocabulary, language learners do not remember words). In this case thesauri and dictionaries of synonyms are useful.
- *Alternate lexemes.* The user is looking for alternative expressions for a particular word or phrase.
- *Vocabulary.* The user desires to master vocabulary of a particular concept for example, feelings, rail travel.

General language dictionaries can assist with language production in several ways providing grammar, collocations and examples in dictionary articles (Frankenberg-Garcia, 2015). However, as several studies have proven, many users are not aware that dictionaries include information on syntactic patterns (Béjoint, 1981). This statement even applies to English university students (Herbst & Stein, 1987). Several studies reveal that the users may encounter problems using the dictionary for writing because of the lack of essential dictionary search skills (Tomaszczyk, 1979; Nesi & Meara, 1994; Atkins & Varantola, 1997). Furthermore, it is often the case that users do not search the whole entry and concentrate only on the first sense (Nuccorini, 1994) or misinterpret the data at all (Nesi & Meara, 1994). Problems may be encountered due to native language interference.

3. 7 Preliminary Conclusion

The ESSCD combines several lexicographic dictionary types. First of all, it is a specialized dictionary, since it is devoted to covering a restricted range of information, namely the language of application forms, as mentioned above. The ESSCD is a learner's dictionary i.e. applicants are learners of English

as a Foreign Language. The ESSCD fulfils functions of an active and passive translation dictionary, since it helps with text encoding and decoding. Last but not least, the ESSCD is an onomasiological dictionary, because it presents words according to concepts (e.g. personal information, skills, education etc.).

Based on the example of *Accounting Dictionaries*, the Function Theory of Lexicography addresses the triadic structure of bilingual specialized dictionaries that consists of (1) a lexicographical database; (2) a user interface where one or more dictionaries are integrated; (3) a search engine that mediates between the database and the user interface (Fuertes-Olivera & Tarp, 2014, p. 195). The current chapter discusses the three structures of the ESSCD.

4.1 The Lexicographical Database of the ESSCD

We should distinguish between the dictionary database and the presentation of dictionary data. The database is a storage system. Kunze and Lemnitzer (2007, p. 12) define lexical databases as „digitale lexikalische Ressourcen, die in einer Form abgespeichert sind, dass die einzelnen Datensätze konsistent im Hinblick auf eine formale Beschreibung ihrer Struktur sind". A single dataset may correspond to a dictionary article or its part. Lexical databases are especially beneficial for lexicographers because some lexicographical data such as example sentences can be related to different articles. In this section, we provide the description of the logical structure of the ESSCD and introduce all the components on the basis of concrete examples.

Markup languages have been developed for the purpose of annotating documents and texts in electronic form. As mentioned earlier, the Extensible Markup Language (XML) was used to annotate the ESSCD. The Text Encoding Initiative (TEI) proposes guidelines for annotating different text types in electronic form based on the XML syntax rules. The TEI Guidelines include a chapter on the representation of dictionaries and lexicographic resources. One of the major advantages of XML is the fact that it is a descriptive markup language, i.e. it describes the role a specific piece of information plays in a text or document, for example: the author of a text can be marked up with the descriptive tag or element <author>; a date that appears in a text can be annotated with the tag <date>; a place name can be marked using the tag <place>. The extensible nature of XML allows for dictionary-specific tags to be created. Documents based on the TEI are composed of two main sections: *TEI header* and *TEI text*. These sections of the ESSCD will be discussed in detail in the following subchapters.

4.1.1 ESSCD Header Database

The lexicographical elements are grouped hierarchically. The TEI Header consists of four parts:

- *File Description <fileDesc>* groups the bibliographic description of the ESSCD.

- *Encoding Description <encodingDesc>* provides the description of the implemented annotation for fields and subfields of the ESSCD.
- *Profile Description <profileDesc>* describes the English-Spanish Comparable Corpus, the languages used and the situation in which the corpus was compiled.
- *Revision Description <revisionDesc>* lists changes and the history of corrections carried out on the ESSCD.

The first part, File Description, is obligatory for all the TEI documents. The File Description section of the ESSCD's header contains:

- *Title Statement <titleStmt>* which presents information about the title of the dictionary and people responsible for the ESSCD's content.
- *Extent <extent>* provides information on how many articles have been written.
- *Publication Statement <publicationStmt>* indicates where the ESSCD is published and its copyright status. The ESSCD is protected under the CC-BY-NC license. It means that the content of the ESSCD is free for non-commercial use. Modification and distribution of ESSCD data is allowed, however, the copyright should be included and the applied changes indicated, the author and the original title should be credited.
- *Source Description <sourceDesc>* provides information about the primary and secondary sources used for the compilation of the ESSCD.

While creating an XML document, first of all, the declaration and the root element should be determined. For the ESSCD the root element designates the type of document (dictionary). The ESSCD's header is the following:

```
<?xml version="1.0" encoding="UTF-8"?>
<dictionary>
 <teiHeader>
    <fileDesc>
        <titleStmt>
            <title>English-Spanish Summer Camp Dictionary</title>
            <author>Olha Morys</author>
            <supervisor>Idalete Dias</supervisor>
            <sponsor>European Master in Lexicography</sponsor>
            <respStmt>
                 <resp>compiled by</resp>
                 <name>Olha Morys</name>
                 <resp>reviewed by</resp>
                 <name>Idalete Dias</name>
            </respStmt>
        </titleStmt>
         <extent>1000 entries</extent>
        <publicationStmt>
            <authority>University of Minho</authority>
            <availability status="free">This is an open access work</availability>
            <licence>CC-BY-NC</licence>
```

```
            <pubPlace> <ref target="http:// ">http://</ref> </pubPlace>
        </publicationStmt>
        <sourceDesc>
            <bibl>
              <corpora>
                <p>The Corpus of Contemporary American English</p>
                <p>iWeb Corpus</p>
                <p>English Web Corpus 2015</p>
              </corpora>
               <dictionaries>Linguee English-Spanish Dictionary</dictionaries>
               <other_sources>
                 <p>International Exchange of North America</p> <p>Indeed</p>
               </other_sources>
            </bibl>
        </sourceDesc>
    </fileDesc>
    <encodingDesc>
        <classDecl>
            <taxonomy>
               <category xml:id="tax.a"><catDesc> Personal information</catDesc> </category>
               <category xml:id="tax.b"><catDesc>Education and Background</catDesc></category>
               <category xml:id="tax.c"><catDesc>Job preference</catDesc></category>
               <category xml:id="tax.d"><catDesc>Work experience</catDesc></category>
               <category xml:id="tax.e"><catDesc>Skills</catDesc>
               <category xml:id="tax.e.1"><catDesc>Camp counselor</catDesc>
                 <subcategory xml:id="tax.e1a">
                    <subcategoryDesc>Adventure</subcategoryDesc></subcategory>
                 <subcategory xml:id="tax.e1b">
                    <subcategoryDesc>Aquatics and Waterfront</subcategoryDesc></subcategory>
                 <subcategory xml:id="tax.e1c">
                    <subcategoryDesc>Arts and Crafts</subcategoryDesc></subcategory>
                  <subcategory xml:id="tax.e1d">
                    <subcategoryDesc>Horse Riding</subcategoryDesc></subcategory>
                  <subcategory xml:id="tax.e1e">
                    <subcategoryDesc>Music</subcategoryDesc></subcategory>
                  <subcategory xml:id="tax.e1f">
                    <subcategoryDesc>Performing Arts</subcategoryDesc></subcategory>
                  <subcategory xml:id="tax.e1g">
                    <subcategoryDesc>Special Skills</subcategoryDesc></subcategory>
                  <subcategory xml:id="tax.e1h">
                     <subcategoryDesc>Sports</subcategoryDesc></subcategory>
                </category>
               <category xml:id="tax.e.2"><catDesc>Support staff</catDesc>
                  <subcategory xml:id="tax.e2a">
                     <subcategoryDesc>Administration</subcategoryDesc></subcategory>
                  <subcategory xml:id="tax.e2b">
                     <subcategoryDesc>Housekeeping</subcategoryDesc></subcategory>
                  <subcategory xml:id="tax.e2c">
                     <subcategoryDesc>Kitchen</subcategoryDesc></subcategory>
                  <subcategory xml:id="tax.e2d">
                     <subcategoryDesc>Maintenance</subcategoryDesc></subcategory>
                  <subcategory xml:id="tax.e2e">
                     <subcategoryDesc>Security</subcategoryDesc></subcategory>
                   </category>
                 </category>
                <category xml:id="tax.f"><catDesc>Reference</catDesc></category>
                <category xml:id="tax.g"><catDesc>Documents</catDesc></category>
                <category xml:id="tax.h"><catDesc>Medical history</catDesc></category>
                <category xml:id="tax.i"><catDesc>Visa information</catDesc></category>
                <category xml:id="tax.j"><catDesc>Emergency contact</catDesc></category>
                </taxonomy>
        </classDecl>
    </encodingDesc>
        <profileDesc>
            <creation>
                <date value="07-2019">July 2019</date>
```

```
88.                 <rs type="city">Braga, Portugal</rs>
89.             </creation>
90.             <langUsage>
91.                 <language id="EN-USA" usage="35%">American English</language>
92.                 <language id="EN-CA" usage="35%">Canadian English</language>
93.                 <language id="ES-MX" usage="20%">Mexican Spanish</language>
94.                 <language id="ES-ES" usage="10%">Spanish of Spain</language>
95.         <note>The information corresponds to the self-compiled corpus</note>
96.             </langUsage>
97.             <settingDesc>
98.                 <p>Texts crawled from the Web</p>
99.          <note>The information corresponds to the self-compiled corpus</note>
100.            </settingDesc>
101.         </profileDesc>
102.         <revisionDesc>
103.            <change>Header finished</change>
104.            <who>Olha</who>
105.            <date>01.07.2019</date>
106.         </revisionDesc>
107.     </teiHeader>
108.</dictionary>
```

The ESSCD header contains the relevant meta-information about the dictionary. Users can become familiar with the ESSCD size (<extent>), its creation date (<date value="07-2019">July 2019</date>) and place (<rs type="city">Braga, Portugal</rs>), terms and conditions under which the dictionary content may be copied (<license>). Moreover, the header provides detailed description of the dictionary base that informs lexicographers where they can find useful information on the elaboration of the dictionary articles and may generate interest in ESSCD sources on part of the user. The *Revision Description* is particularly important for the lexicography team since it indicates what changes were made, when and by whom. Thus, each collaborator has access to information on the latest modifications The *Encoding Description* with categories and subcategories predefines the access structures of the ESSCD. The ESSCD allows for the inclusion of attributes which designate the properties of the element, for example, <availability status="free">, where <availability> is an element, *status is* the attribute name with its value "free" placed in double quotes.

4.1.2 ESSCD Text Database

The TEI Text section contains the tagged dictionary content.

The TEI Guidelines propose the following 3 sections for representation of texts in digital format:

- Front matter <front>
- Body of text <body>
- Back matter <back>

As can be seen from the proposed elements, the TEI primarily focuses on print dictionaries. Compared to paper-based dictionaries, the online ones do not display articles in a linear order. Therefore, the

concepts of *front and back matter* do not apply to online dictionaries. Consequently, new elements have to be introduced to cater for the specificities of online lexicographic resources. Since online dictionaries do not distinguish between front and back matter, the element <outer_texts> was created to annotate the corresponding texts in the current dictionary project. The creation of own project tags followed the XML naming convention: if the tag name consists of two words, they can be linked by an underscore separator. In this way we have distinguished the elements created specifically for the project from the TEI elements in which the first letter of the second word of the element appears in capital letters and without spacing between the first and second word that make up the tag. The ESSCD database also contains empty elements. The empty elements only have the closing tag, e.g.
 to markup a line break and <hi/> to annotate information that is being highlighted.

The text of the ESSCD consists of:

- Outer texts[60] <outer_texts>
- Body of text <body>

The body contains all the dictionary articles. Each article has its own ID number and a section specifying the domain to which the lemma belongs. The ID number helps with the internal organization of the database and is not displayed for users. Each *Collocation Group* is organized by patterns according to the part of speech of the collocation base (verb, noun, adjective). Each pattern section contains a list of English collocations and the corresponding Spanish translations. We distinguish between translations of the headword itself <cit type="translation_lemma"> and the translation of collocations and corpus examples <cit type="translation" xml:lang="es">. If the article includes more than one part of speech, e.g. noun and verb in the articles *dance*, *cook*, the Grammar Group sections <gramGrp> appear as subsections of the supertag <word_class>. The majority of the elements used in the project's XML schema are presented in the article for the lemma *kayaking*:

```
1.   <?xml version="1.0" encoding="UTF-8"?>
2.  <!DOCTYPE entry SYSTEM "entry.dtd">
3.  <entry id="7">
4.      <form type="lemma">
5.          <orth>kayaking</orth>
6.          <pron notation="ipa"> <pc>/</pc>'kaɪækɪŋ <pc>/</pc>
7.           <ptr type="audioFile" target="kayaking.mp3" title="listen to pronounciation"/></pron>
8.      </form>
9.      <gramGrp>
10.         <pos>noun</pos>
11.     </gramGrp>
12.     <domain>
13.         <usg dom="#tax.e1">Camp Counselor</usg>
```

[60] The components of outer texts are discussed in Section 4.2.3 Outer Texts.

```
        <usg subdom="#tax.e1b">Aquatics and Waterfront</usg>
    </domain>
    <sense>
        <cit type="translation_lemma" xml:lang="es">
            <gramGrp>
                <pos>sustantivo</pos>
                <gen>masculino</gen>
            </gramGrp>
            <quote>kayak</quote> <quote>piragüismo</quote>
        </cit>
    </sense>
    <collocGrp>
        <pattern id="verb">
            <colloc n="1">
                <cit xml:lang="en">
                    <quote>go kayaking</quote>
                </cit>
                <cit type="translation" xml:lang="es">
                    <quote>pasear en kayak</quote>
                </cit>
            </colloc>
        <pattern id="noun">
            <colloc n="2">
                <cit xml:lang="en">
                    <quote>kayaking trip</quote>
                </cit>
                <cit type="translation" xml:lang="es">
                    <quote>viaje en kayak</quote>
                </cit>
            </colloc>
            <colloc n="3">
                <cit xml:lang="en">
                    <quote>kayaking tour</quote>
                </cit>
                <cit type="translation" xml:lang="es">
                    <quote>excursión en kayak</quote>
                </cit>
            </colloc>
            <colloc n="4">
                <cit xml:lang="en">
                    <quote>whitewater kayaking</quote>
                </cit>
                <cit type="translation" xml:lang="es">
                    <quote>kayak en aguas rápidas</quote>
                </cit>
            </colloc>
            <colloc n="5">
                <cit xml:lang="en">
                    <quote>sea kayaking</quote>
                </cit>
                <cit type="translation" xml:lang="es">
                    <quote>kayak de mar</quote>
                </cit>
            </colloc>
            <colloc n="6">
                <cit xml:lang="en">
                    <quote>ocean kayaking</quote>
                </cit>
                <cit type="translation" xml:lang="es">
                    <quote>kayak oceánico</quote>
                </cit>
            </colloc>
        </pattern>
            <pattern id="adjective">
                <colloc n="7">
                    <cit xml:lang="en">
```

```
                        <quote> guided kayaking </quote>
                    </cit>
                    <cit type="translation" xml:lang="es">
                        <quote>kayak guiado</quote>
                    </cit>
                </colloc>
            </pattern>
        <pattern id="preposition">
            <colloc n="8">
                <cit xml:lang="en">
                    <quote>kayaking along</quote>
                </cit>
                <cit type="translation" xml:lang="es">
                    <quote>kayak a lo largo de</quote>
                </cit>
            </colloc>
        </pattern>
    </collocGrp>
    <corpusExamples>
        <example n="1">
            <cit xml:lang="en">
                <quote>Did you normally wear a lifejacket when you would <phrase rend="italics
">go kayaking</phrase>?</quote>
                <bibl>
                   <title>Mystery on the Hudson; Accident or Murder?</title> <date>2015</date>
                 </bibl>
            </cit>
            <cit type="translation" xml:lang="es">
             <quote>¿Usaba normalmente un chaleco salvavidas cuando iba a hacer kayak?</quote>
            </cit>
        </example>
        <example n="2">
            <cit xml:lang="en">
                <quote>Later in the day I got a chance to <phrase rend="italics">try out kayak
ing</phrase>.</quote>
            </cit>
            <cit type="translation" xml:lang="es">
                <quote>Más tarde en el día tuve la oportunidad de probar kayak.</quote>
            </cit>
        </example>
        <example n="3">
            <cit xml:lang="en">
                <quote>The best way to learn <phrase rend="italics">whitewater kayaking </phra
se>is with a good instructor.</quote>
                <bibl>
                  <author>Beth Geiger</author><title>Rapid racers</title><date>2002</date>
                </bibl>
            </cit>
            <cit type="translation" xml:lang="es">
                <quote>La mejor manera de aprender kayak en aguas rápidas es con un buen instr
uctor.</quote>
            </cit>
        </example>
    </corpusExamples>
    <note type="General tasks of a kayaking instructor">
        <unit>
            <cit xml:lang="en">
                <quote>to teach proper paddling techniques</quote>
            </cit>
            <cit xml:lang="es">
                <quote>enseñar técnicas adecuadas de remo</quote>
            </cit>
        </unit>
        <unit>
            <cit xml:lang="en">
                <quote>to explain boating safety</quote>
```

```
            </cit>
            <cit xml:lang="es">
                <quote>explicar la seguridad de la navegación</quote>
            </cit>
        </unit>
        <unit>
            <cit xml:lang="en">
                <quote>to provide assistance to campers using the equipment</quote>
            </cit>
            <cit xml:lang="es">
                <quote>asistir a los campistas que utilizan el equipo</quote>
            </cit>
        </unit>
    </note>
</entry>
```

In order to ensure that all articles are well-formed and follow the standardization principles, a Document Type Definition (DTD) was created. The DTD allows us to define the elements used to annotate metadata and the dictionary content: (1) the structure and content of the dictionary; (2) the order in which these elements must appear; (3) if the elements are mandatory or optional; (4) which attributes are used with the respective elements. The DTD contains the rules that must be followed during the annotation process.

As can be seen from the example below, the content model defines the content of each entry as follows: <!ELEMENT entry (form, gramGrp*, domain, sense*, wordClass*, collocGrp*, corpusExamples*, note*, image*)>. The root element <entry> is composed of the following information units: <form> is composed of orthography and pronunciation; <gramGrp> includes elements part-of-speech, subcategorization, gender and tense; <domain> specifies which (sub)field the headword belongs; <sense> contains the translation equivalent(s) in Spanish with grammatical information on the translation equivalent; <collocGrp> contains collocations according to their patterns, e.g. <pattern id="noun"> for *whitewater kayaking*; <corpusExamples> includes examples that explain the usage of collocations; <note> consists of encyclopedic information; <image> indicates the title and the copyright of the image used to represent the concept. Once all the elements are determined, we can move to the declaration of the attributes. Attributes are declared for an element in a sequence/list form. Each attribute declaration contains the element, the attribute associated with the element, the type of value admitted by the attribute and its status as optional or mandatory. The attribute declaration <!ATTLIST cit xml:lang (en | es) #REQUIRED> specifies that the element <cit> admits the mandatory (#REQUIRED) attribute xml:lang and its value is either "en" or "es" (English or Spanish). The DTD of the ESSCD articles is an external DTD:

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT entry (form, gramGrp*, domain, sense*, wordClass*, collocGrp*, corpusExamples*, note*, image*)>
<!ELEMENT id (#PCDATA) >
<!ELEMENT form (orth,pron)>
<!ELEMENT orth (#PCDATA)>
<!ELEMENT pron (#PCDATA |pc|ptr)*>
<!ELEMENT pc (#PCDATA)>
<!ELEMENT ptr EMPTY>
<!ELEMENT gramGrp (pos, subc*, gen*, tns*)>
<!ELEMENT pos (#PCDATA)>
<!ELEMENT subc (#PCDATA)>
<!ELEMENT gen (#PCDATA)>
<!ELEMENT tns (#PCDATA) >
<!ELEMENT domain (usg+) >
<!ELEMENT usg (#PCDATA)>
<!ELEMENT sense (cit)>
<!ELEMENT cit (gramGrp|quote |synset| bibl|usg)*>
<!ELEMENT quote (#PCDATA | phrase)*>
<!ELEMENT phrase ( #PCDATA)>
<!ELEMENT synset (quote+) >
<!ELEMENT bibl ( author| title| date)*>
<!ELEMENT author (#PCDATA)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT date (#PCDATA)>
<!ELEMENT wordClass (gramGrp,sense,collocGrp,corpusExamples)>
<!ELEMENT collocGrp (pattern*, colloc*)>
<!ELEMENT pattern (colloc+)>
<!ELEMENT colloc (cit+)>
<!ELEMENT corpusExamples (example+)>
<!ELEMENT example (cit+)>
<!ELEMENT note (unit+)>
<!ELEMENT unit (cit| source)*>
<!ELEMENT source (#PCDATA |ref)*>
<!ELEMENT ref (#PCDATA)>
<!ELEMENT image (imageDesc, ptr)>
<!ELEMENT imageDesc (imageTitle, copyright)>
<!ELEMENT imageTitle (#PCDATA)>
<!ELEMENT copyright (#PCDATA)>
<!ATTLIST entry id CDATA #REQUIRED >
<!ATTLIST form type (lemma| headword) #REQUIRED>
<!ATTLIST pron notation CDATA #REQUIRED>
<!ATTLIST ptr type (audioFile|image) #IMPLIED>
<!ATTLIST ptr target CDATA #IMPLIED >
<!ATTLIST ptr title CDATA #IMPLIED >
<!ATTLIST usg dom CDATA #IMPLIED >
<!ATTLIST usg subdom CDATA #IMPLIED >
<!ATTLIST cit type (translation |translation_lemma) #IMPLIED>
<!ATTLIST cit xml:lang (en |es) #REQUIRED >
<!ATTLIST pattern id (noun | verb | adjective|adverb|preposition) #REQUIRED>
<!ATTLIST colloc n CDATA #IMPLIED >
<!ATTLIST example n CDATA #IMPLIED >
<!ATTLIST phrase rend (italics) #REQUIRED >
<!ATTLIST note type CDATA #IMPLIED >
<!ATTLIST tns type (past_simple|past_participle)  #REQUIRED>
```

4.2 The User Interface

When a query result appears on the screen, “the data ceases to be a part of the database and becomes a dictionary” (Nielsen & Almind, 2011, p. 143). One of the most important tasks of the lexicographer is to decide where particular data should be located. In terms of data presentation, the

lexicographer needs to bear in mind the reference skills of the potential users: "[i]deally, an online dictionary interface will combine simplicity (for those who cannot be bothered) with sophistication (for those who can). A reasonable way to achieve this is to offer a simple default interface with an optional advanced alternative" (Lew, 2013a, p. 29). Taking into consideration the general theory of lexicography, the section *User Interface* will describe the macrostructure, microstructure and outer texts of the ESSCD, despite the fact that there is no clear separation between microstructure, macrostructure and outer texts in electronic dictionaries (Engelberg & Lemnitzer, 2009, p. 166). In contrast to HTML, XML does not specify how the content should be displayed, i.e. XML being a descriptive markup language does not specify how text should be formatted or displayed, instead it describes the role the units of information play in the text. In the current project the styling language XSLT (eXtensible Stylesheet Language) can be used to format the text.

4.2.1 Macrostructure

Normally, dictionaries imply semasiological order for the macrostructure, however such organization does not reflect conceptually grouped vocabulary. Therefore, the ESSCD will offer a different thematic macrostructure that better reflects the semantic arrangement of the vocabulary. According to the application documents of users, the ESSCD's macrostructure is composed of 10 conceptual fields, which are subdivided into other conceptual subfields. The macrostructure of the ESSCD will include around 1000 articles. In the frame of this dissertation, seven dictionary articles, namely *kayaking*, *camping*, *child care*, *arts and crafts*, *cook*, *guitar* and *dance* from the field "Skills" were elaborated.

4.2.2 Microstructure

Microstructure refers to the hierarchical internal structure of a dictionary article. The microstructure of online dictionaries clearly differs from the microstructure of print ones, i.e. the user can select whether to display or hide specific article items (Engelberg & Lemnitzer, 2009). Engelberg and Lemnitzer differentiate between the real and virtual microstructure. The latter appears directly on the screen. The former includes all microstructural items that are available for a dictionary article. Data in the ESSCD has to be presented in a maximally comprehensible form, therefore the main task is to find a way to present dictionary articles so that the target audience can retrieve information as fast as possible. The study of Tono (2010) revealed that if the microstructure differs from the users' expectations, they became confused and perform a slow search. Lew's study (2004) with Polish learners of English as a Foreign Language has proved that a microstructure, which is too rich, might cause information overload. When the user is overwhelmed with information, this leads to *information stress* (Bergenholtz

& Bothma, 2011, p. 55). Hence, in the ESSCD we use menus to enable users to select the item providing the data that they need. The *headword level* contains information on spelling, pronunciation, grammatical characteristics of the headword and its translation. The *collocations level* includes all the collocations grouped by patterns (noun, verb, adjective etc. as collocates). The *example level* illustrates collocation entities within corpus examples. The last *level corresponding to the notes section* provides encyclopedic information related to the headword. Furthermore, the ESSCD should allow the users to choose the interface language, either English or Spanish. The ESSCD incorporates typographical structural indicators, i.e. it styles lemmata in bold, collocations within examples in italics, makes use of font types and colors.

4.2.3 Outer Texts

In terms of printed dictionaries, "[d]ictionary users are known to allocate little time to the study of these prefatory matters" (Busane, 1990, p. 28). Similarly, the first two international studies on online dictionary use prove that the detailed information on searching and navigating is not essential for most users. Only 4,4% of informants reported that the presence of a clear introduction is important (Muller-Spitzer & Koplenig, 2014). The components beyond the dictionary articles are known as outer texts in printed dictionaries. To distinguish commonly accepted *outer texts* from their print counterparts, Klosa and Gouws (2015, p. 148) suggest the term *outer features* for online dictionaries since "the elements presented in the outer domain of online dictionaries do not all belong to the broad category of texts". With regard to the ESSCD's genuine purpose and its functions, the dictionary will provide the following outer features:

- *User guidelines* will be found under the link "About the project" that provides information on the dictionary content and search functions. The ESSCD should incorporate hyperlinks from the user guidelines to the corresponding entries. The ESSCD will also include question mark symbols in places where questions may arise.
- *Grammar* will be incorporated as mouse-over effect. If users do not understand what is for example a *mass noun*, they will place the mouse over the grammar information and a brief explanation will appear.
- *Questions and Answers*
- *User profile,* where users can adjust the user interface according to their needs, add dictionary articles to the list or play some language games, i. e. the ESSCD fulfills edutainment functions.

Special attention has been paid to the outer text "Questions and Answers" which consists of:

- *Types of camps*. For the users to prepare themselves for a stay in a summer camp, it is essential to know in what type of camp they are going to work. The majority of camps are traditional ones. However, many applicants are placed in other camp types such as religious camps, girl guide camps, wilderness camps, underprivileged camps (also called camps for disadvantaged children), special needs camps (also called camps for children with physical or mental disabilities), agency summer camps, day camps etc. (for the brief camp description see Appendix 14).
- *List of the documents for the J-1 visa**. This section contains an asterisk because the politics of the US department of State may change. The obligatory documents for all the nationalities to obtain this visa type are DS-2019 Form and SEVIS[61] Receipt. The former document is issued by a visa sponsor that verifies the participant's eligibility to take part in cultural exchange programs, summer camp programs among them.
- *Paperwork at camp*. Once the applicants arrive at the camp, they need to obtain a Social Security Number and fill in further forms, e.g. W-4 which is the tax form. By the way, applicants can claim their taxes back, as long as they are not US citizens.
- *What should I pack?* The applicants should have their passport with visa and the DS-2019 Form, otherwise they will not enter the USA. In terms of suitcase packing, we recommend applicants take the following: copies of all documents, pairs of underwear and socks, warm sweater, bathing suits, Flip Flops, t-shirts, trainers or walking shoes, jeans, rain coat, toothbrush and toothpaste, deodorant. These recommendations are very useful for young people who go to the camp for the first time.
- *Pre-departure orientation* should contain presentations or videos about the potential cultural differences and camp life in detail.

The *Questions and Answers* section will also provide answers to questions which users have asked in a forum.

4.2.3 Search Options and Navigation structures. Access Structures of the ESSCD

The search engine helps users to retrieve the data from the database. The ESSCD should provide four search options. The macrostructure is an elementary access structure in the dictionary (Kunze & Lemnitzer, 2007), so users can navigate the dictionary using mainly a conceptually-oriented access

[61] SEVIS stands for Student Exchange Visitor Information System, where the US Department of State controls the process of all J-1 visa applicants.

structure (see Figure 16). In addition, the ESSCD should offer a semasiological A-Z search. The index-based search will provide clickable letter sections. Users need to click on the initial letter and then further navigate the target letter section. The third search option enables users to type their query directly in the search field. The ESSCD search engine should include the feature corresponding to automatic term completion from an index. This function becomes active after the user has typed in a minimum of four characters. Fuzzy-spelling search will be implemented as the *Did you mean?* technique, i. e. the search results will show suggestions for a query that was misspelled. The last search option of the ESSCD should incorporate *internal and external cross-referencing.* Thanks to the former feature any word within an entry may be looked up in the dictionary immediately. Such cross-references such as synonyms, antonyms or superordinate terms are considered to be helpful for the users. Internal references will not only refer to other entries, but also lead to dictionary components outside the dictionary articles. For example, if a user would like to know what a *noun phrase* is, he/she can access the corresponding grammatical information with just one click. Moreover, each article of the ESSCD includes the field(s) and subfield(s) to which the lemma belongs to. By clicking on the respective (sub)field, the user has access to a list of all the lemmata belonging to the same (sub)field.

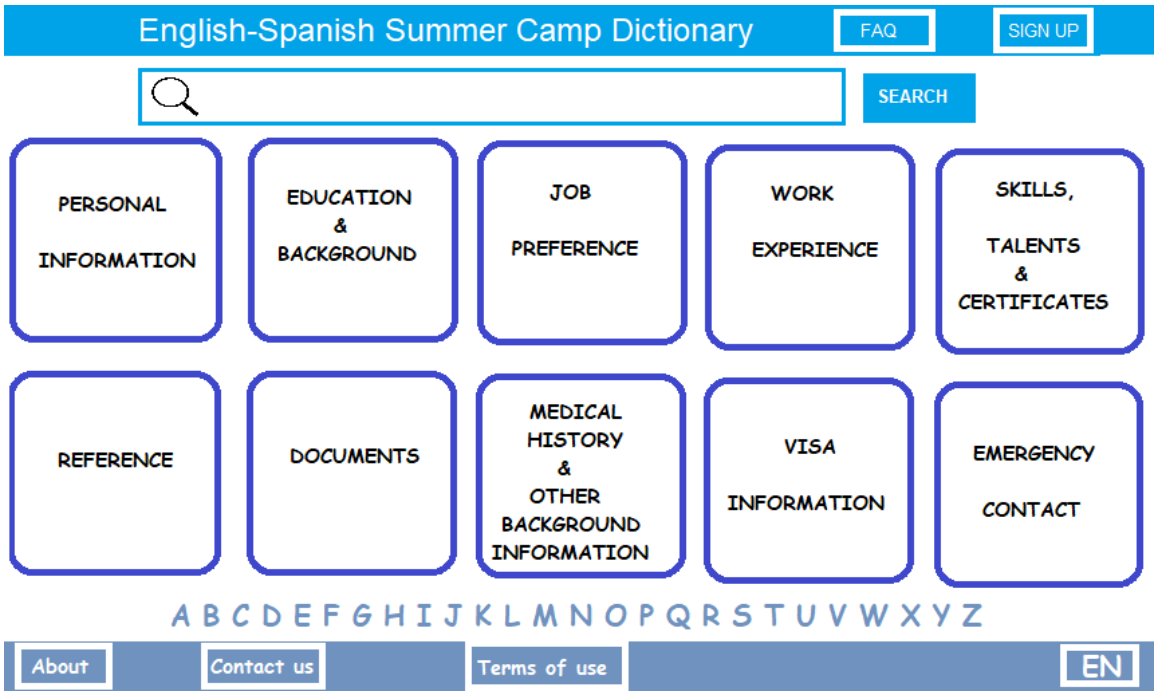


Figure 16: Draft of the ESSCD start page

4.3 Preliminary Conclusion

The ESSCD database has been annotated with the Markup Language XML developed to represent texts and documents in electronic format. The Text Encoding Initiative (TEI) Guidelines follow the syntactic rules of XML and propose tags for the electronic encoding of a wide variety of text types including dictionaries. Chapter 9 of the Guidelines proposes tags for encoding lexicographic information. However, the TEI Guidelines for lexicographic resources focus primarily on the encoding of print dictionaries. This means that the TEI Guidelines present some lacunae when it comes to encoding born digital lexicographic materials and products. Therefore, during the database creation process, new lexicographical elements were included into "the relational hierarchy" (Nielsen & Almind, 2011, p. 143). During the encoding process some elements were modified or added. In order to save consultation time, the ESSCD's microstructure should be organized clearly and provide various search options.

When it comes to the user needs, Bergenholtz and Tarp (1995, p. 24) indicate the following required types of information to support *foreign-language text production in specialized dictionaries*:

(1) orthography, gender, pronunciation, irregularity, collocations, usage information;

(2) standard label (DIN, ISO etc.), field label or brief explanation.

Taking into consideration the *translation from foreign language*, the following information types will be needed (Bergenholtz & Tarp, 1995, p. 24):

(1) on the foreign language: word class, gender, pronunciation, collocations, irregularity;

(2) on the native language: orthography, gender, pronunciation, irregularity, collocations, usage information.

The ESSCD follows the suggestions of Bergenholtz and Tarp (1995)[62] but also takes into account the ESSCD user profile and their specific needs. Since the ESSCD is a monodirectional dictionary, it is unlikely that the Spanish native speaker will need information on Spanish pronunciation. This section describes the entry items of the ESSCD and its organization. The main task is to select which information categories to be included. According to Wiegand, Feinauer and Gouws (2013, p. 343), a dictionary article consists of two structural components, called comments, namely a *comment on form* and a *comment on semantics*.

5.1 Comment on Form

The comment section on the form of the ESSCD articles includes orthography, pronunciation and grammatical data on the word class and inflection.

5.1.1 Orthography

The spelling variants of the lemma, such as *go-cart* and *go-kart*; *tie-dye* and *tye dye, tennis racket* and *tennis racquet* etc., are included into the ESSCD:

```
<form type="lemma">
   <orth>go-cart</orth> or <orth>go-kart</orth>
</form>
```

The spelling form which is proposed by the IENA application appears first followed by the spelling variant. Spelling variants sometimes belong to different varieties of English language. An example is *counselor* for American English and *counsellor* for British English. The ESSCD includes American forms

[62] More recent research on the Function Theory of Lexicography (Fuertes-Olivera & Tarp, 2014, p. 101) proposes the following data to be included into the dictionary article of online specialized dictionaries: collocations/word combinations; example sentences; contextual definitions and background text.

as lemmata and British ones as spelling variants. The British spelling variants have corresponding indications:

```
<form type="lemma">
   <orth>couselor</orth>
   <form>
      <usg type="geo">UK</usg>
      <orth>counsellor</orth>
   </form>
   <pron notation="ipa"> <pc>/</pc>ˈkau̯n(t)-s(ə-)lər <pc>/</pc> <ptr type="audioFile"
target="counselor" title="listen to pronunciation"/></pron>
</form>
```

5.1.2 Pronunciation Data

The ESSCD includes both phonetic transcription according to IPA and the audio representation. Lew (2013) determines two principal benefits of keeping the graphic transcription in online dictionaries. The first one is clarity i.e. the language learners are unlikely to perceive all the phonological details at once as they do it from the phonetic system of the mother tongue. Another significant advantage is the indexical function e.g. IPA transcriptions are based on commonly accepted criteria/rules. As can be seen from the element <form> above, the phonetic transcription in the ESSCD is placed between slashes. The audio pronunciation of the ESSCD should be recorded by native English speakers. The pronunciation as a data category is included into the ESSCD because it serves for text production purposes during the interview with recruiters and camp directors.

5.1.3 Grammatical Data

Grammatical data of the ESSCD contains, as a rule, the word class and verb inflection. Since many nouns and verbs in the English language have the same orthographic representation, the word class also designates the meaning, thus contributing to text reception. Most of the dictionaries encode concomitant metalinguistic information such as abbreviations, codes or symbols. This causes comprehension problems among the users.

The aim of the ESSCD is to avoid contracted forms as the ones above-mentioned and to make the dictionary use as transparent as possible. Therefore, the ESSCD offers worded grammar, indicating the part-of-speech in its full form e.g. “noun” instead of “n”. Recent studies with the participation of 117 Dutch and 606 Polish learners of English (Bogaards & van der Kloot, 2002; Dziemianko, 2012) also reveal that the majority of informants prefer worded grammar patterns.

Some lemmata contain subcategorization information. For example, *wakeboarding* is a mass noun, *next of kin* – a noun phrase, *arts and crafts* – a plural nominal word combination:

```
<gramGrp>
   <pos>noun</pos>
   <subc>mass</subc>
</gramGrp>
```

The majority of the ESSCD lemmata are nouns[63] as the dictionary is organized conceptually. However, if the lemma sign for a verb and for a noun is the same, the ESSCD provides explicit word forms for verb, for example *tie-dye*:

```
<gramGrp>
   <pos>verb</pos>
   <tns type="present_simple">tie-dyes</tns>
   <tns type="present_participle">tie-dyeing</tns>
   <tns type="past_simple">tie-dyed</tns>
   <tns type="past_participle">tie-dyed</tns>
</gramGrp>
```

Bergenholtz and Tarp (1995, p. 51) recommend that "specialised dictionaries intended for translation from the foreign language into the native language should provide, in addition to relevant collocations, a minimum of **grammatical information on the native language**"[64]. The ESSCD includes the word class and the gender of Spanish equivalents.

5.2 Comment on Semantics

The comment section on semantics of the ESSCD articles consists of translation equivalents, collocations, examples and notes, which will be discussed in the following sections. The mentioned data categories help users to decode the meaning of the lemma and indicate its usage.

5.2.1 Translation Equivalents

The selection of equivalents for lemmata with cross-cultural differences poses a serious challenge for lexicographers. Concepts in one language do not always have their counterparts in another. Bergenholtz and Tarp (1995) propose two solutions for the lack of target language equivalents: paraphrase and partial equivalence. Regarding the latter option, the difference between the lemma and partial equivalent must be explained. The ESSCD also contains headwords which have zero correspondence in

[63] L' Homme (2003) claims that specialized dictionaries usually contain nouns that define activities or processes and leave out the corresponding verbs.

[64] Emphasis in original.

the Spanish language. *Next of Kin's address* is an example. The term *Next of Kin* in the USA is defined legally and refers to (1) the nearest blood relatives of a person who has died, including the surviving spouse; (2) anyone who would receive a portion of the estate by the laws of descent and distribution if there were no will[65]. *Next of Kin* can be paraphrased as *familiar más cercano*[66]. In terms of the ESSCD *Next of Kin's address* that belongs to the field *Emergency Contact* refers to the address of the ***nearest living relative***. This ESSCD article should include explanatory information.

Searching for the right equivalent of the collocation *guitar riff*, *acordes de guitarra* is one of the translation variants in the corpus examples of the online dictionary *Linguee*. As Bergenholtz and Tarp (1995, p. 27) claim, encyclopedic information is needed "to distinguish between the degree of equivalence". In fact, *guitar riff* is a broader term than *guitar chords*. *Guitar riff* means a short repeated tune[67], usually a solo improvisation[68] and includes a series of chords. Therefore, *riff de guitarra* is a better translation option. Another potential difficulty may arise when the user lacks encyclopedic information on what a *riff* actually is. Therefore, the collocation should provide a cross-reference to the article *riff,* which should contain this encyclopedic information essential for laypeople. As Bergenholtz and Tarp (1995) highlight, linguistic and encyclopedic information supplement each other.

5.2.1.1 Corpora for Translation

This section aims at identifying Spanish-English translation candidates from a self-compiled comparable corpus. As a rule, parallel corpora are used in bilingual lexicography. However, there are more texts available in one language than those translated by professional translators. Only few web pages of summer camps have both, English and Spanish versions, therefore we decided to compile a comparable corpus. Basically, comparable corpora are primarily used for three main purposes[69]: (1) extraction of bilingual dictionaries by identifying candidate translation equivalents; (2) as a source to train Statistical Machine Translation (SMT); (3) learning purposes, i.e. exploring a particular type of text in both languages prior to engaging in translation. The current thesis aims at identifying candidate translation equivalents. In order to extract translation pairs from a comparable corpus, two principal approaches are used: (1) seedwords[70], "w1 is a candidate translation of w2 if they tend to co-occur [within a particular window] with the same seed words"; (2) lexico-syntactic patterns, for example

[65] *Next of Kin*. In *Legal Dictionary*. Retrieved from https://legal-dictionary.thefreedictionary.com/next+of+kin

[66] Closest relatives.

[67] *Riff*. In *Collins Dictionary*. Retrieved from https://www.collinsdictionary.com/dictionary/english/riff

[68] *Riff*. In *Merriam-Webster Dictionary*. Retrieved from https://www.merriam-webster.com/dictionary/riff

[69] The use options 1 and 2 are suggested in *Building and Using Comparable Corpora* (Sharoff et al. (Eds.), 2013), the 3rd use option of comparable corpora is proposed by Aston (1999).

[70] Also called *vector space approach*.

NOUN + bar, for example *chocolate bar*. Aston (1999) suggests the use of comparable corpora in the creation of terminology banks by comparing candidate lists of terms in two languages. Considering Aston's proposal, the process of identifying the translation equivalents in the self-compiled corpus was organized as follows.

After single word and multi-word terms are extracted, similar text fragments containing these terms should be identified. The next step is to analyze the detected text fragments and interpret the results. Additional lexicographic resources need to be consulted when a lemma occurs less frequently in the corpus. Since the Spanish part of the corpus is considerably smaller, we decided that the translation process will run from the Spanish to the English language because this increases the possibility of finding translation equivalents.

In order to provide an overview of identified term translations and highlight specialized terms, Table 5 was created. The first column of the table contains the single word or multi-word terms identified by Sketch Engine. The second column lists the identified translation equivalent in the English subcorpus. The third and fourth columns present the frequency of the term in the Spanish subcorpus and a reference corpus of the Spanish language, namely *Spanish Web 2011*[71], correspondingly. Like the Spanish subcorpus, the Spanish Web 2011 is a web-based corpus which represents the European and American varieties of the Spanish language and contains 9 497 213 009 words. The fifth and the sixth columns correspond to the frequency of equivalent candidates in the English subcorpus and in the reference corpora for the English language. COCA and the BNC were selected as reference corpora for English in order to identify some term differences between British and American English, if any. If a translation equivalent was not found in the English subcorpus, it is indicated with a dash.

Spanish	***English***	***F1 in Spanish subcorpus of self-compiled corpus***	***Reference Corpus Frequency1***	***F2 in English subcorpus of self-compiled corpus***	***COCA Frequency 2***	***BNC Frequency 2***
acampadores	campers	117	35	4 115	1 701	54
acampante	camper	63	1 594	1 830	1 176	97
equitación	riding	123	9 670	29	16 682	2 470
piragua	canoe	77	4 030	426	3 041	733
hípica	horse riding	78	4 736	9	28	60

[71] Spanish Web Corpus 2011. In Sketch Engine. Retrieved from https://www.sketchengine.eu/estenten-spanish-corpus/

acampar	camp	222	40 842	15 362	39 971	4 885
campista	camper	60	3 190	1 830	1 167	97
tirolina	zip line	41	1 091	13	58	0
kayak	kayak	109	23 890	38	1 489	117
doma	-	60	9 400	-	-	-
senderismo	hiking	87	20 997	111	6 149	72
albergue	shelter	319	102 855	33	12 675	2 150
kitesurf	-	33	2 194	-	-	-
surfear	-	33	4 046	-	-	-
ecuestre	-	56	15 676	-	-	-
navegar	navigate	425	185 008	81	3 715	242
rapel	abseiling/ rappelling	32	5 152	2/ 1	1/9	84/ 0
aventura	adventure	674	340 229	485	10 670	1 989
galopes	-	20	14	-	-	-
animación	entertainment	325	165 138	2 206/ 33	53 287/ 20 723	22 831/ 2 232
compañerismo	fellowship	61	24 877	52	3 519	896
convivencia	living together/ coexistence	27	294 928	5/ 4	997/ 823	158/ 102
canoa	canoe	49	19 387	595	3 041	733
escalada	climbing	126	78 280	145	10 547	2 146
salvaguardar	safeguard	83	59 313	24	1 879	1 384
vivencial	experiential	33	15 042	154	1 371	139
animador	entertainer	68	53 906	2	1 468	252
chapuzón	plunge	15	4 554	6	3 152	1 667
jinete	-	52	42 190	-	-	-
litera	bunk bed	18	8 609	3	152	32
remar	paddle	33	28 226	55	2 266	661
campismo	camping	10	2 926	557	4 228	431
cabaña	cabin	69	85 398	765	11 708	1 268
patinaje	skating	19	16 926	13	3 890	203
cuatrimotos	-	7	161	-	-	-
lavandería	laundry	30	35 815	98	3 890	537

pulsera	bracelet	28	33 059	5	2 580	286
remo	paddle	27	35 102	55	1 467	661
consejeros	counselors	26	33 668	267	8 898	355[72]
sabbath	-	9	5 431	-	-	-
travesía	crossing	41	71 542	62	10 845	2 505
socorrista	lifeguard	12	4 797	16	786	87
interiorizar	internalize	6	4 524	46	649	121

Table 5: Single word term equivalents from the self-compiled corpus

As can be observed from the results in Table 5, several single word terms in the Spanish language have their multi-word equivalents in English, for example *bunk bed* for *litera*, *horse riding* for *hípica*. Considering the small size of the self-compiled corpus, the occurrence of terms in it, e.g. *campers*, *camp*, is in 5 times higher than in general language corpora.

The keyword *zip line* was not found in the BNC corpus even though it does not belong to a particular variety of English. Looking at the corpus results for *abseiling* (BNC – 84; COCA – 1) and *rappelling* (BNC – 0; COCA – 9), it can be concluded that the former term seems to belong to British and the latter to American English. Thanks to the extraction of the single word terms, new equivalents for the term *camper* were discovered. While the *Linguee English-Spanish Dictionary* only suggests the Spanish equivalent *campista,* apart from this term, two more options were found in the self-compiled corpus, namely *acampador* and *acampante*.

The Spanish keywords *piragua* and *canoa* find their equivalent of *canoe* in the English subcorpus. Some water clubs suggest that *piragua* is the general term for what we know today as the "kayak" and the "canoe"[73]. Similarly, Wikipedia explains, "pirogue [...] is any of various small boats, particularly dugouts and native canoes"[74]. Another term that captured our attention is *Sabbath* which refers to the day of the week kept by some religious groups for rest and worship[75]. There are nearly 600 religious summer camps in the USA[76]. Applicants who are followers of the Christian religion but go to work for a summer camp for followers of another religion should be aware of the specific beliefs and practices of this religion, for example an applicant for a Jewish Summer Camp must necessarily be acquainted with

[72] The American variant *counselor* was not found in the BNC, the indicated frequency refers to the British spelling variant.

[73] See La diferencia entre un kayak y una canoa y una piragua. In Kayakismo.com. Todo sobre el kayak. Retrieved from http://www.kayakismo.com/tipos/diferencia-kayak-canoa-piragua.html

[74] See *Pirogue*. In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Pirogue

[75] *The Sabbath*. In *Cambridge Dictionary*. Retrieved from https://dictionary.cambridge.org/dictionary/english/sabbath

[76] According to the search results in American Camp Association.

the Jewish Sabbath. Moreover, participants who are hired as kitchen staff should be informed that Jewish people do not mix meat with dairy products in their meals.

From the extracted keywords it can also be concluded that the notion of camps in Spain and America is slightly different. For example, the campers in the Spanish corpus sleep in *albergue* (hostel), whereas in the English subcorpus campers stay in *cabins* and share them with *cabin mates*, among them we find a *cabin leader* and all together they form a *cabin group* or *cabin family*. It should be noted that the listed keywords with "cabin" have a high mutual information score: e.g. *cabin leader* - 8,5, *cabin mates* -10,8.

An interesting single word term found in the English part of the self-compiled corpus is *chapel*. Analyzing the contexts of this term, we have observed that in non-religious camps it refers to an event, when a camp community gets together and reflects on its values. *Chapel* also involves thanksgiving, singing, telling stories etc.: "[t]he ageless traditions of Sunday soap baths, hand-written letters to parents, chapel talks and special dinners. Chapel is a tradition as old as Pathfinder, with a community gathering and personal time for reflection and appreciation"[77]. On the web sites of religious summer camps, *Chapel* is spelled with a capital letter and refers to a small church of the camp. Another interesting term is *tetherball*, that refers to a ball game for two players, whose aim is to wrap the ball around the pole[78].

Spanish	***English***	***F1***	***Reference Corpus Frequency1***	***F2***	***COCA frequency 2***	***BNC Frequency2***
campamento de verano	summer camp	513	163	2 823	1 040	23
pensión completa	full board/ room and board	128	233	1/ 7	47/ 456	166/ 7
reparto de habitación	-	121	1	-	-	-
comida picnic	picnic lunch	317	1	3	141	28
viaje de fin	weekend trip	136	62	1	116	19
fiesta de despedida	-	130	51	-	-	-
profesor acompañante	-	69	8	-	-	-

[77] Camp Pathfinder. See in English-Spanish Comparable Corpus sources

[78] See *Tetherball*. In *Collins Dictionary*. Retrieved from https://www.collinsdictionary.com/dictionary/english/tetherball

rango de edad	age range	99	169	11	1 384	229
bicicleta de montaña	mountain bike	58	173	12	1 053	186
vehículo de apoyo	-	55	31	-	-	-
preferencia de dieta/ restricciones alimenticias/ restricciones alimentarias	dietary restrictions	38/ 2/ 3	0/ 111/ 104	19	78	3
actividad física	physical activity	35	3 082	444	6 403	113
turno de campamento	camp session	30	1	106	6	0
entorno natural	natural environment	133	836	102	852	183
espíritu de servicio	spirit of service	25	132	1	18	0
actividad acuática	water activity	14	96	4	20	4
país de origen	home country	13	1 453	27	704	142
equipo de salto	-	12	6	-	-	-

Table 6: Multi-word term equivalents from the self-compiled corpus

Looking at Table 6, the following observation on translation equivalents can be made: many NOUN + NOUN collocations where the collocate describes the base in English are translated with a construction NOUN + PREPOSITION + NOUN (e.g. *river kayaking - kayak de río*, *whitewater kayaking - kayak en aguas rápidas*, *weekend camping - acampada de fin de semana*, *overnight camping - acampada de noche*). As such collocation constructions are quite common for the English language, it is important to include them in the dictionary. *Fiesta de despedida* which is known as *farewell party* found several equivalents in the English subcorpus of the self-compiled corpus. Some camps usually organize a *closing ceremony* or *closing celebration* others *closing campfire* at the end of every session.

The extraction of multi-word terms once again reveals that some terms, such as *camp session* (is not presented in the BNC at all), the term *summer camp* occurs more frequently in the self-compiled corpus than in general language corpora. Looking at the Spanish keyword *turno de campamento*, where *turno* usually refers to a *shift*, the first thought went to *day camp* and *overnight camp* types, which operate during specific hours/shift. Having analyzed the context of the term, we came to the conclusion that *turno de campamento* means *camp session*. Overall, the multi-word keyword and terms

table once again demonstrates the difference between American and British English, for example, *room and board* vs. *full board* and illustrates the presence of this variety in the self-compiled corpus.

5.2.2 Collocations

You shall know a word by the company it keeps.

Firth, 1957, p. 11

There is no commonly-agreed definition for the term "collocation". In this thesis we use it to refer also to *multi-word keywords and terms* that were extracted from the self-compiled corpus. Regarding the collocation structure, Hausmann distinguishes between the base and the collocate: "Die *Basis* ist ein Wort, das ohne Kontext definiert, gelernt und übersetzt werden kann. Der *Kollokator* ist ein Wort, das beim Formulieren in Abhängigkeit von der Basis gewählt wird und das folglich nicht ohne die Basis definiert, gelernt und übersetzt werden kann (Hausmann, 2007, p. 218).

Bilingual dictionaries are criticized for including redundant information, which users do not need, as well as for the lack of collocations (Atkins, 1996). Mastering collocations is one of the greatest challenges in second language learning. From the point of view of dictionary research on collocations, various studies regarding collocation search and users' capability of applying the information in production tasks were carried out (e.g. Laufer & Waldman, 2011; Nesselhauf, 2005). In particular, their results revealed that more than 50% of the errors are caused by the influence of the native language, i.e are interlingual. Therefore, one of the main purposes of the ESSCD is to provide an exhaustive collocation list and correct collocation equivalents. The main selection criteria for collocations was their relevance to the subject of summer camps, thus *child care credit*, *parental child care* are not relevant terms in summer camp domain.

The corpora which we have searched, enrich each other. For example, for the lemma *kayaking*, the English Web 2015 proposes *go kayaking*. On the other hand, the COCA corpus suggests types of kayaking depending on its venue e. g. *sea kayaking*, *ocean kayaking*, *river kayaking*, *whitewater kayaking*. General language corpora offer a wide range of collocates, therefore we had to select which ones are relevant for the ESSCD.

5.2.2.1 Collocations from the Self-compiled Corpus

Various statistical approaches and association measures determine collocations within the corpus. Corpus collocation extraction goes back to the 1980s (Sinclair, 1991). However, at that time the corpus size 10-50 million tokens was too small to retrieve useful examples (Geyken & Lemnitzer, 2016, pp. 221-222). How large should a corpus be to enable the extraction of representative collocations? 85

million words were used for the *Oxford Learner's Dictionary of Academic English*[79]. Another problem that we still have to tackle today is insufficient accuracy. The above-mentioned term "collocation" is set in a wide framework and the extraction tools frequently obtain "weak" word combinations such as *a new shirt* (Geyken & Lemnitzer, 2016, p. 222). Therefore, it is helpful to use association measures. Keywords and terms extracted from the self-compiled corpus with higher mutual information[80] are likely to better represent the summer camp domain, e.g. *activity specialists* (MI 8,7) vs. *intake specialist* (MI 14), *age group* (MI 5,5) vs. *cabin group* (MI 5,7). *Age group* is a common collocation, while *cabin group* relates directly to summer camp life.

Here are some interesting collocations extracted from the English subcorpus: *sleeping bag*, *intersecting chords*, *tennis court*, *canoe trip*, *day session*, *camp season*, *staff member*, *family camp*, *camp spirit*, *camp experience*, *leadership skills*, *camp staff*, *senior staff*, *former camper*, *personal growth*, *camp placement*, etc. The last term refers to the hiring process, i.e. if the candidate is hired, it means that he/she is placed in a camp. An illustrative word combination is *hacky sacks* that refers to a game played with footbag, which is filled with sand, rice or other materials. An interesting example taken from the output of the candidate term extraction process is *t-shirts tanks*. Even though the expression itself does not exist, it leads us to two synonyms *sleeveless shirt* and *tank top*, "named after tank suits, one-piece bathing suits of the 1920s worn in tanks or swimming pools"[81]. An additional incorrect compound is *shield cage.* Used separately, *hockey shield* or *hockey cage* refer to a helmet for hockey. In nearly every camp, campers go canoeing and kayaking. For their personal belongings not to get wet, they use *dry bags*, a type of container that protects things from getting wet. An interesting multi-word expression found in the English subcorpus is *staff-to-camper ratio* that refers to the number of counselors divided by the number of campers, e.g. 1:2. A *camp trunk* or *foot locker* are primarily used to store the belongings and "become a dresser, stepstool, seat and table for [...] camper"[82]. Another term found in the self-compiled corpus, *in loco parent* is a legal term meaning "in place of the parent": "[w]hile state laws vary, camp professionals generally serve in loco parentis (in place of the parent)" (see American Camp Association). An additional term that is worth mentioning is *seasonal staff*. Only few staff members work in camp for the whole year. The majority of camp staff is seasonal, i.e. works during the camp season. One more identified keyword is a *polar bear swim/ polar bear dip* that refer to "early

[79] *Oxford Learner's Dictionary of Academic English*. Retrieved from https://www.oxforddictionaries.com/oldae

[80] Mutual information is computed in Cygwin terminal using the script *ngrams.py*.

[81] *Tank top*. In *The Free Dictionary*. Retrieved from http://www.thefreedictionary.com/tank+top

[82] Camp Pathfinder. See in English-Spanish Comparable Corpus Sources.

morning swims, a camp activity where you jump in the lake in the early morning before breakfast"[83]; "a refreshing optional dip before each breakfast"[84].

Lists of keywords and terms extracted from the English subcorpus were transformed into a network (see Figure 17) using Gephi software[85]. Once collocations were extracted, the data was formatted in one CSV file for the importation to Gephi's "data laboratory" (see Tables 7, 8). Gephi constructs a network composed of edges and nodes. Nodes are collocation bases and edges are connections between the base and its collocate. At the beginning of the process, the information in the graph appears in random order. Therefore, the Modularity Algorithm implemented in Gephi was applied. It attributes nodes to clusters of highly interconnected nodes. In Figure 17, few collocation pairs are "isolated" such as *dining hall*, *sleeping bag*, while the majority belong to clusters, composed of pairs in which one collocation element is connected to other word(s). Part-of-speech (PoS) category was included into the database because the network can be further elaborated with adjectives and verb collocations. Other categories, e.g. association or statistical measures like mutual information or log-likelihood can be employed.

ID	Label	PoS	Frequency
camp_NN	Camp	NN	15 326
summer_NN	Summer	NN	4 008
staff_NN	Staff	NN	3 818

Table 7: Collocation data for import to Gephi as nodes table[86]

Source	Target	Type	Frequency
summer_NN	camp_NN	Direct	1 270
camp_NN	staff_NN	Direct	457
camp_NN	employment_NN	Direct	260

Table 8: Collocation data for import to Gephi as edges table

The next network, *SkillNet*, illustrates collocates of the base *skill*. The graph is based on the results of the self-compiled corpus and includes 34 nodes (see Figures 18 and 19). Each node has its value, which corresponds to the frequency of the word in the corpus. The edges' value equals the frequency with which the base appears with its collocate. It was possible to include more nodes, however the graph would look overloaded. The color scheme was applied to distinguish between word classes: light

[83] NYQUEST Camp Canada. See in English-Spanish Comparable Corpus Sources.

[84] Camp Pathfinder. See in English-Spanish Comparable Corpus Sources.

[85] Brett (2017) uses Gephi in the paper *Collocate networks in the language of crime journalism*.

[86] See Appendices 15, 16 for the full database of collocation networks.

green for adjectives (*outdoor skills*), light blue for verbs (*to master skills*, *to demonstrate skills*) and red for nouns (*skill acquisition*, *communication skills*). As can be observed from Figures 18 and 20, the thicker the edge is, the more frequent the words occur together, for example, *skill development*, *leadership skills*, *social skills*. The *Fruchterman Reingold* layout algorithm and the *Expansion* layout algorithm have been used to make the appearance of *SkillNet* user-friendly for retrieval. Some properties of the mentioned layouts, namely the speed and scale factor have been manipulated before running the algorithms. The edges were ranked by weight. The nodes were ranked by degree and value. We consider the use of such visualizations particularly important in terms of the current study. They not only inform about the typical collocations of summer camps but also demonstrate the strength between the base and its collocate as well as emphasize the connection between collocations.

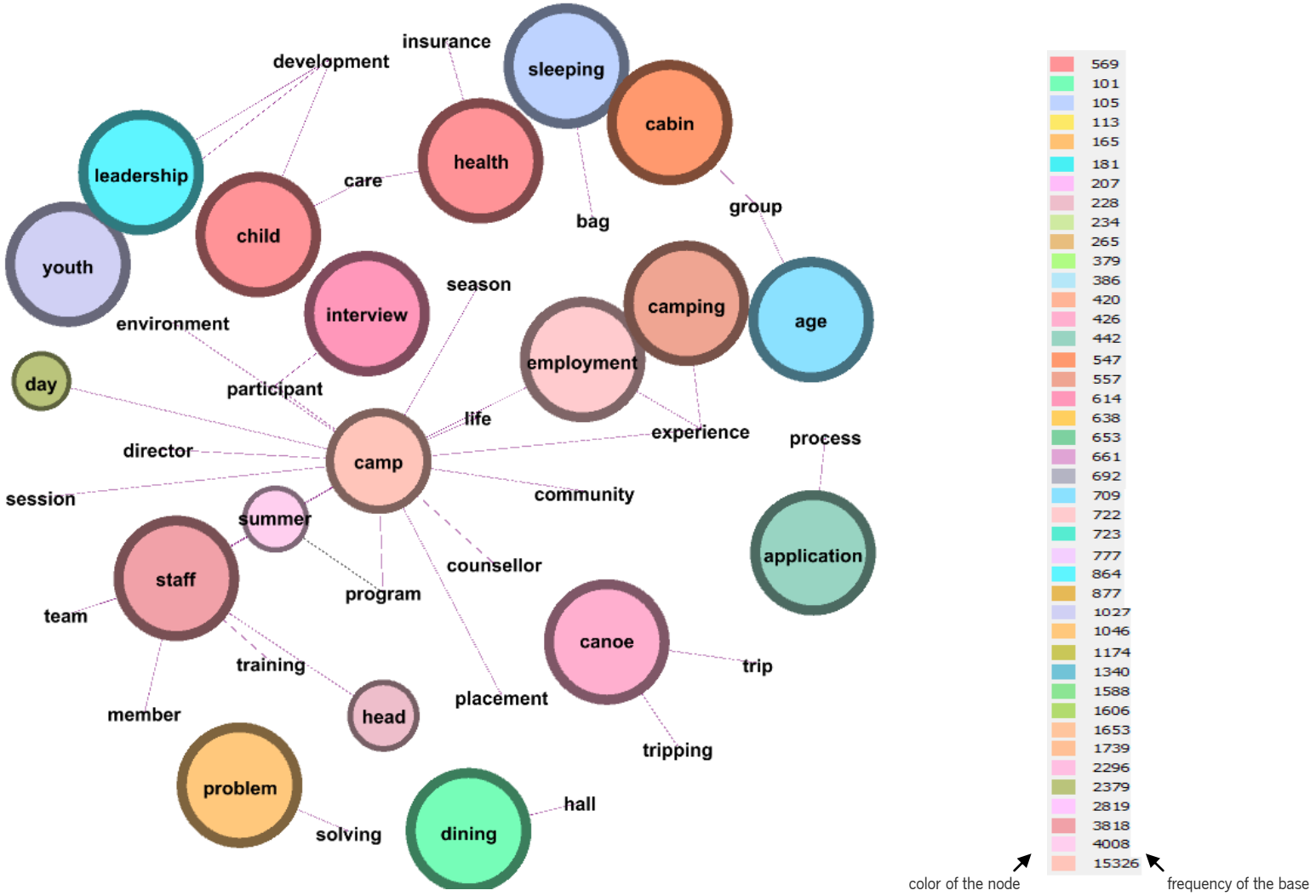


Figure 17: Bigrams extracted from the English part of the self-compiled corpus with frequency 35 and above*.

* Networks in Figures 17, 18, 19 and 20 were created using Gephi by the author of the thesis.

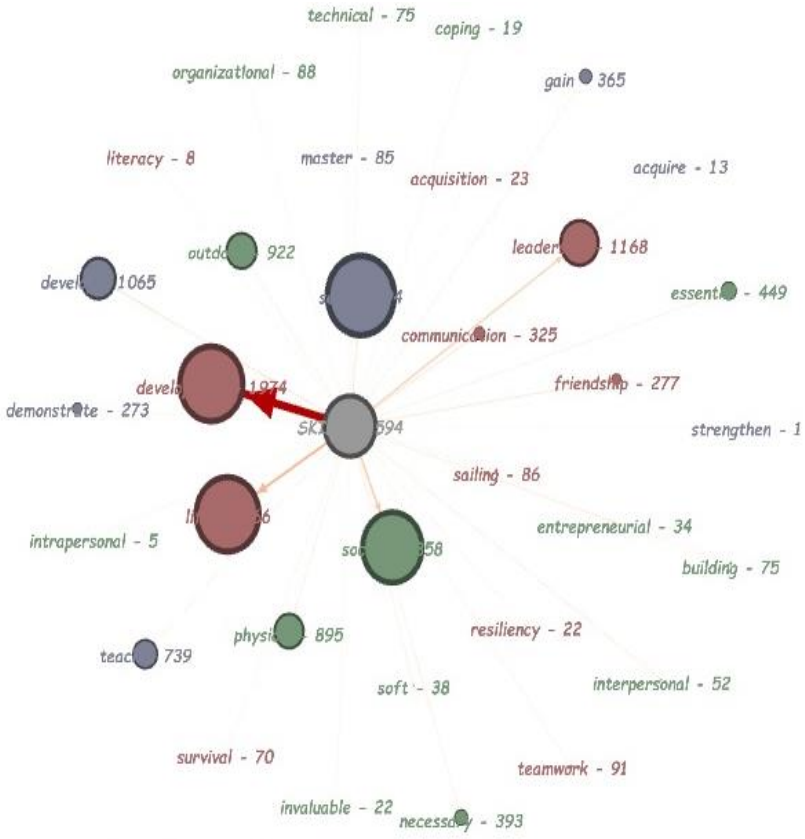


Figure 18: SkillNet

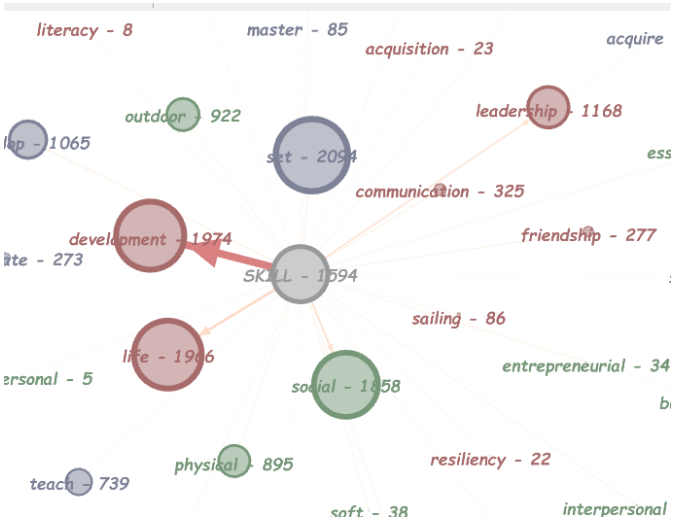


Figure 19: Expansion of the central part of the SkillNet

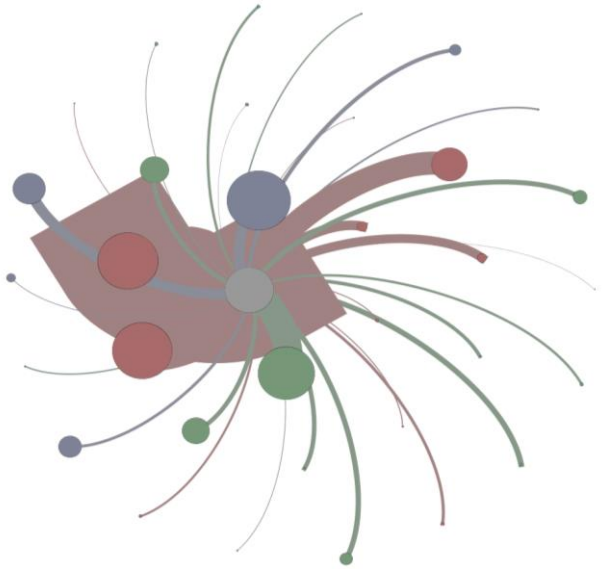


Figure 20: SkillNet. Edges visualization.

5.2.3 Example Sentences

5.2.3.1 Corpora vs. Lexicographer-made Examples

It is commonly assumed that large corpora are the most representative sources for dictionary examples because they include the language of native speakers and illustrate real life communication, i.e. are authentic. On the one hand, lexicographer-made examples are criticized as isolated and overloaded with information (Fox, 1987). On the other hand, lexicographers have linguistic intuition and are also native speakers of a particular language. Furthermore, the "artificial" examples might be preferred in learner's dictionaries (Laufer, 1992). Some studies have proven that neither language teachers nor native speakers were able to distinguish between "natural and made-up examples" (Rundell & Maingay, 1990). Similarly, Laufer (1992) examined whether corpus examples contribute to better comprehension by learners compared to example sentences written by lexicographers. The experiment has shown that lexicographers' examples are remarkably more helpful to understand new words and slightly more useful for production purposes. The advocate of corpus examples is Sinclair, who also describes the role of corpora for the compilation of the *Collins Cobuild Dictionary* (1987). Other lexicographers may state that the lexicographers' examples are focused on the usage point (sense item) that should be illustrated and authentic examples only confuse the user (Engelberg & Lemnitzer, 2009, p. 236). Rundell (2010) suggests linking corpora to online dictionaries[87]. On the one hand, corpora allow users to search authentic examples. On the other hand, it could lead to information overload as the task of the lexicographer is to process the corpus data, evaluate it and select the best examples. Moreover, the target group of the ESSCD is not skilled at corpus searching. As Bothma (2011) and Tarp (2011) claim, data should be repackaged and adopted to the specific information needs.

5.2.3.2 Corpus Examples in the ESSCD

ESSCD's examples aim to assist users with text production and collocation use. We have followed four main principles in selecting corpus examples. First of all, they need to illustrate the use of collocations in a sentence. Secondly, if the example is too long, the part of the example which does not contain the collocation is omitted in a way that the sense of the sentence is not lost. We believe that too long examples increase information costs[88]. As a rule, examples contain bibliographic information with author, source and date. However, not all concordances in corpora are fully annotated with resource details. Moreover, web-based corpora include only links to the web pages which often cannot be accessed as

[87] *Base lexicale du français* offers access to such an external database and gives the user the possibility to select the corpus. For more about the integration of online dictionaries and corpora, see Heid, Prinsloo and Bothma (2012).

[88] Nielsen and Fuertes-Olivera (2013) distinguish between *search-related information costs* (the efforts which the user makes to look up the needed information) and *comprehension-related information costs* (the efforts the user makes to understand the data found).

the web site was removed. In these cases, the ESSCD does not include bibliographic information. Last but not least, examples should preferably contain encyclopedic information. For instance, for the article *guitar:*

- *Then, as you **tune the guitar up** for the first time, give the strings a good tug to help stretch them in.*
- *Until one learns to **tune** their **guitar** manually they can always use the principle of beats to tune their stringed instrument.*
- *One day he picked up an old **classical guitar** he'd bought at a garage sale, **tuned it up**, and played a few chords.*

The examples give instructions on how to prepare a guitar before playing and explain the differences between guitar types. Ideally examples describe camp life:

- *Campers will participate in **arts and crafts**, sports, organized games, field trips and much more.*

By default 5 examples will be shown, users have an option to click on the plus sign, thereby expanding the listed examples.

5.2.4 The Notes Section

The notes section usually contains the general tasks and duties of a particular working position, for example, those of a dance instructor. Moreover, notes sometimes include idioms and the equipment related to a certain activity such as *camping supplies*. The note tab might offer *action verbs*, e.g. for the lemma *cook*, it lists *stir*, *simmer*, *chop*, etc. Notes can be also designed for error prevention clarifying the use of easily-confused words. Encyclopedic information will be given in the notes section as well. The *Did you know that?* feature will offer interesting etymological facts of some entries, for example, *lacrosse* goes back to Canadian French *la crosse*, that means the *crooked stick*[89]. The term *go-carts* underwent a semantic change from the baby walker of the 17th century to nowadays a small racing car[90].

5.3 Preliminary Conclusions

The dictionary is an utility tool, hence the main aim of the ESSCD is to ensure that the user obtains information in a straightforward way: "[th]e perfect dictionary is one in which you can find the thing you are looking for preferably in the very first place you look" (Haas, 1962, p. 48). The ESSCD contains data that respond to the needs of the users: a word's pronunciation, its spelling, basic grammatical

[89] See *Lacrosse*. In *Merriam-Webster Dictionary*. Retrieved from https://www.merriam-webster.com/dictionary/lacrosse

[90] See *Go-cart*. In *Online Etymology Dictionary*. Retrieved from https://www.etymonline.com/word/go-cart

information, syntactic behavior in a sentence and collocational restrictions. In sum, the ESSCD offers accurate and complete lexicographic data and allows the users to retrieve the exact required information that they need[91]. A good dictionary shall cover the most frequent collocations and illustrative examples with the collocations (Laufer & Waldman, 2011). Examples are selected carefully and contribute to a better understanding of the meaning of the word and illustrate collocational behavior within the sentence.

[91] Most lexicographers welcome the possibility of showing exactly the relevant information categories in a particular lookup situation, no less and no more, tailored to the specific needs and skills of the user (Trap-Jensen, 2010, p. 1142).

CHAPTER 6 GENERAL CONCLUSION

6.1 Limitations

The current project has certainly several limitations and some aspects can be improved. One of them is the corpus size. Due to the time constraints, we limited the information extraction process to 38 camp websites out of the more than 3 600 available camps. During the extraction of the webpage URLs, external links, for example, to foundations, Facebook or Twitter camp pages appeared. This led to some noise in the corpus, fact that reduced the number of representative examples. Another aspect that can be developed is the insertion of labels for regional variety of American and Canadian English. The dictionary compilation requires an immense input of effort and time. It goes without saying that the compilation of the ESSCD should involve experts with subject-field knowledge, skilled and experienced translators, experts with high level of proficiency in the source (English) and target (Spanish) languages, web page programmers and designers. Given that the current lexicographic project is the result of the work developed as part of the Master's thesis, subject to a limited time span, it was not possible to compile a dictionary manual. Since the ESSCD is not a finished reference work, we have not conducted any research on dictionary use and could not conclude whether the users' needs are satisfied. It is implied that the indended users of the ESSCD should assess and evaluate the dictionary.

6.2 Discussion on Future Directions

Lemmata of dictionary articles that were written in the frame of the current thesis did not pose a serious comprehension problem, hence no definitions were included and the translation equivalents were sufficient for the understanding of the meaning of the headwords. However, the users are likely to lack assistance to understand particular terms, for example *tetherball*, *tye-dye*, *lacrosse*, *doughboy*[92]. Therefore, such headwords should contain the explanation of meaning in a *definition*. We suppose that this data category should be included into the article structure.

One of the most important future tasks will be to implement novel technologies which will allow users to optimize consultation time. Several lexicographers have already suggested several options. In 1996, Atkins (1996, pp. 13-14) foresees that bilingual dictionaries of the future might operate two *modes of information*. The *equivalence mode* assists the user in the performance of "specific tasks such as translation, comprehension or self-expression" and *contrast mode* "offers ways of contrasting the meaning and syntactic behaviour of chosen words across languages". Since every user has individual

[92] Also called *stick biscuit*.

needs and skills, it might be beneficial to let users decide which data categories should be displayed. Lew (2013a) proposes two possible approaches in terms of entry presentation for online dictionaries. The first is *user-controlled customization* when the user selects lexicographic items or chooses the type of task. The second approach, *application-controlled customization* is based on on-line monitoring of user activities and analysis of their behavior. Thus, application-controlled customization identifies the probable task type that the user is dealing with and settles the presentation mode correspondingly. Similarly to Lew's application-controlled customization, Bothma (2011) talks about *user profiling*, when the system traces users' search behavior and retrieval actions and, as a result, builds the users' profile, which can be saved in the system and implemented for further requests. In fact, Google has already applied this technique. Studying and living in different countries, we have noticed that Google searches are slightly different. This means that Google calculates that the users from Ukraine may prefer certain sources, and the users from Portugal will opt for others. Trap-Jensen (2010) also suggests the implementation of user profiles and two versions, *short version*, which contains basic information needed for text reception and a *long version* with all the available information. Fuertes-Olivera and Tarp (2014) implement four search options in *Accounting Dictionaries*: users can select a reception, translation, knowledge, phrases and expressions tab.

Industries tend to develop steadily. Summer camps are not an exception. One of the placement agencies, *Camp Leaders*, offers applicants a possibility to complete their application via a mobile application. Following this example, it seems feasible to convert the English-Spanish Summer Camp Dictionary into an application with an interactive interface or integrate it into the application of the above-mentioned placement agency. One should keep in mind that the compilation and publishing process of the ESSCD should not take years, because users, their needs, technologies and languages themselves are undergoing a rapid change: "[t]here has never been a more exciting time to produce a new dictionary. Everything is changing, diversifying, and expanding: the English language itself, the technology that helps us to describe it, and the needs and goals of those learning and teaching English" (Rundell, 2002, p. x).

6.3. Conclusion

In the current dissertation we have elaborated on the methodology and steps involved in the compilation of the English-Spanish Summer Camp Dictionary. The ESSCD guides the users through their application process and assists the participants in text reception and text production tasks. General corpora and a specialized self-compiled comparable corpus, additional dictionaries and search

engines were used to build the dictionary base. To be more precise, we extracted single word and multi-word terms in English and Spanish from the self-compiled comparable corpus and determined the translation equivalents. Moreover, the self-compiled corpus served to extract collocations, which were visualized in networks. The main focus of the thesis was to define the structures of the ESSCD and the data categories of a dictionary article. Since e-lexicography allows customization, one of the future tasks will be the creation of the system that offers individualized content according to the specific user needs and situations. It is possible to implement this thanks to user profiles that allow the user to choose presentation mode and information types and also define the default settings. As Lew (2013a, p. 16) points out, "[s]uccessful dictionary use requires two ingredients: (1) high-quality, user-friendly dictionaries and (2) dictionary users who know what they are doing". Therefore the ESSCD should also offer educational videos on dictionary use.

The current dissertation is definitely not the last word on the subject of online onomasiological dictionaries. With the technological rush, online dictionaries are continuously developing and opening up new possibilities for both the lexicographers and end users.

References

English-Spanish Comparable Corpus sources

List of web pages for the English subcorpus

American Camp Association. Retrieved from https://www.acacamps.org/

Camp Canada by NyQuest. Retrieved from https://www.go-nyquest.com/

Camp Davern. Retrieved from https://campdavern.com/

Camp Ekon. Retrieved from http://www.ekon.ca/

Camp Gesher. Retrieved from https://campgesher.com/

Camp Massad. Retrieved from https://campmassad.ca/

Camp MINI-YO-WE. Retrieved from https://www.miniyowe.com/

Camp NeeKauNis. Retrieved from https://www.neekaunis.org/

Camp Northway - Wendigo. Retrieved from https://www.campnorthway.com/

Camp Pathfinder. Retrieved from http://www.camppathfinder.com/

Camp Walden. Retrieved from https://www.campwalden.ca/

Camp White Pine. Retrieved from https://campwhitepine.com/

Canadian Camping Assosiation. Retrieved from http://www.ccamping.org/

Easterseals. Retrieved from http://www.easterseals.com/

Keewaydin. Retrieved from https://keewaydin.org/

Manitou Camp. Retrieved from https://www.manitoucamp.com/

Mathcamp. Retrieved from https://www.mathcamp.org/

Pioneer Camp. Retrieved from https://www.pioneercampmanitoba.ca/

RKY Camp. Retrieved from https://www.rkycamp.org/

Safari Zoo Camp. Retrieved from http://www.zoocamp.com/

YMCA of Greater Halifax/Dartmouth. Retrieved from http://www.ymcahfx.ca/big-cove-ymca-camp/

YMCA Wanakita. Retrieved from https://www.ymcahbb.ca/Wanakita/Home

List of web pages for the Spanish subcorpus

Asociación Mexicana de Campamentos. Retrieved from https://www.campamentosdemexico.com/

Camp Santa Ursula. Retrieved from http://www.campsantaursula.com

Camp Tomacoco. Retrieved from http://www.tomacoco.com.mx/

Camp Valle Verde. Retrieved from http://valleverde.com.mx/

Campamento El Molino. Retrieved from http://www.elmolinogto.com/index.html

Campamento Náutico Bilingüe. Retrieved from https://www.campamentosveranoplaya.es/

Campamentos. Retrieved from https://solocampamentos.com/

Campus Experience. Retrieved from https://campusexperiencermf.com/

Club Hipico Miracampos. Retrieved from https://hipicamiracampos.com/

Encantaría. Retrieved from https://www.encantaria.com

Icaros Campamentos. Retrieved from https://www.icaroscampamento.com/

IKAL Campamento. Retrieved from http://www.ikal.com.mx/default.aspx

Irish Summer Camp. Retrieved from http://www.irishsummercamp.com/

Kin Camp. Retrieved from https://kincamp.com/

Kobal Camp. Retrieved from https://kobal.camp/

Natuaventura. Retrieved from https://natuaventura.com/

Pipiol. Retrieved from https://www.pipiol.com/es/index.php

Planeta Palombina. Retrieved from http://planetapalombina.com/

Rancho Providencia. Retrieved from http://ranchoprovidencia.mx/ecocamp-providencia/

Rockcamp. Retrieved from https://www.rockcamp.es/

Corpora

Corpus of Contemporary American English. Retrieved from https://www.english-corpora.org/coca/

English Web Corpus 2015. Retrieved from https://www.sketchengine.eu/

iWeb Corpus. Retrieved from https://www.english-corpora.org/iweb/

Dictionaries

Accounting Dictionaries = Nielsen, S., Mourier, L. & Bergenholtz, H. (2012). *Accounting Dictionaries*. Odense. Retrieved from: Ordbogen.com

Berufenet = Bundesagentur für Arbeit (BA). (2019). *BERUFENET*. Retrieved from https://berufenet.arbeitsagentur.de/berufenet

Collins Cobuild Dictionary = Sinclair, J. (Ed.). (1987). *Collins Cobuild English Language Dictionary*. London: HarperCollins.

Computer Glossary = *Computer Glossary*. Newton: TechTarget. Retrieved from https://whatis.techtarget.com/

Deutsches Wörterbuch = Grimm, J. & Grimm W. (1995). *Deutsches Wörterbuch.* 33 vol., München.

Diccionario de la Lengua Española = *Diccionario de la Lengua Española*. Madrid: Real Academia Española. Retrieved from www.rae.es

DoL = Hartmann, R.R.K., & James, G. (1998). *Dictionary of Lexicography*. London, New York. Routledge.

DOT = Information Technology Associates. (1995-2011). *Dictionary of Occupational Titles*. Retrieved from https://occupationalinfo.org/

LEO = *LEO*. München: LEO GmbH. Retrieved from https://www.leo.org/

Lexin = Malmgrem, S.-G. (Ed.). (2012). *Lexin*. Retrieved from https://lexin.nada.kth.se/lexin/

Macmillan Open Dictionary = *Macmillan Open Dictionary*. London: Macmillan Publishers. Retrieved from https://www.macmillandictionary.com/open-dictionary/

Merriam-Webster Open Dictionary = *The Open Dictionary*. Spriengfield: Merriam-Webster. Retrieved from http://nws.merriam-webster.com/opendictionary/

OALD 4 = Hornby, A.S., & Cowie, A. P. (Eds.). (1989). *Oxford Advanced Learner's Dictionary of Current English* (4th ed.). Oxford: Oxford University Press.

OpenThesaurus = *OpenThesaurus, Synonyme und Assoziationen.* Potsdam: LanguageTooler GmbH. Retrieved from https://www.openthesaurus.de/

Oxford English Dictionary = Murray, J. A. H. et al. (Eds.). (1st edn 1884-1928, Supplement 1933 and 1972-1986). *The Oxford English Dictionary*. Oxford: Clarendon Press.

Wiktionary = *Wiktionary, the free dictionary*. San Francisco: Wikimedia Foundation. Retrieved from https://www.wiktionary.org/

Other Literature

Abel, A., & Meyer, Ch. M. (2016). Wörterbuchbenutzungsforschung. In A. Klosa & C. Müller-Spitzer (Eds.), *Internetlexikografie. Ein Kompendium* (pp. 249-290). Berlin, Boston: de Gruyter.

Aston, G. (1999). Corpus use and learning to translate. *Textus*, 12, 289-314. Retrieved from https://www.sslmit.unibo.it/~guy/textus.htm

Atkins, S. (1996). Bilingual Dictionaries: Past, Present and Future. In M. Gellerstam, J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström, and C. R. Papmehl (Eds.), *EURALEX'96 Proceedings* (pp. 515–546). Göteborg: Department of Swedish, Göteborg University. Retrieved from https://euralex.org/publications/bilingual-dictionaries-past-present-and-future/

Atkins, S. (1996). Bilingual Dictionaries: Past, Present and Future. In M. Gellerstam, J. Jarborg, S.-G. Malmgren, K. Norén, L. Rogstrom, & C. Rojder Papmehl (Eds.), *Euralex '96 Proceedings* (pp. 515-546). Göteborg: Göteborg University. Retrieved from http://www.euralex.org/elx_proceedings/Lexicography%20and%20Natural%20Language%20Pr

ocessing/B.T.S.%20Atkins%20-%20Bilingual%20Dictionaries%20Past,%20Present%20and%20Future.pdf

Atkins, S., & Knowles, F. (1990). Interim report on the EURALEX/ AILA research project into dictionary use. In T. Magey and J. Zigany (Eds.). *BudaLex '88 Proceedings* (381-392). Budapest: Akadémiai Kiadó.

Atkins, S., & Varantola K. (1997). Monitoring dictionary use. *International Journal of Lexicography,* 10 (1), 1-45.

Atkins, S., Clear, J., & Ostler, N. (1992). Corpus Design Criteria. *Literary and Linguistic Computing*, 7 (1), 1-16. Retrieved from http://www.natcorp.ox.ac.uk/archive/vault/tgaw02.pdf

Béjoint, H. (1979). The use of informants in dictionary making. In R. R. K. Hartmann (Ed.). *Dictionaries and their Users: Proceedings of the 1978 BAAL Seminar on Lexicography* (25-30). Exeter: University of Exeter Press.

Béjoint, H. (1981). The foreign student's use of monolingual English dictionaries: A study of language needs and reference skills. *Applied Linguistics*, 2 (3), 207-222.

Béjoint, H. (1994). *Tradition and Innovation in Modern English Dictionaries*. Oxford: Clarendon Press.

Béjoint, H., & Moulin, A. (1987). The place of the dictionary in an EFL programme. In A. Cowie (Ed.). *The dictionary and the language learner* (pp. 381-392). Tübingen: Niemeyer.

Bergenholtz, H., & Bothma, T. (2011). Needs-adapted Data Presentation in e-Information Tools. *Lexikos*, 21, 53-77.

Bergenholtz, H., & Bothma, T. (2011). Needs-adapted Data Presentation in e-Information Tools. *Lexikos*, 21, 53-77.

Bergenholtz, H., & Gouws, R. (2010). A New Perspective on the Access Process. *Hermes*, 44, 103-127. Retrieved from http://pure.au.dk/ws/files/34579091/hermes_44_bergenholtz_gouws_1_.pdf

Bergenholtz, H., & Tarp S. (Eds.). (1995). *Manual of Specialised Lexicography. The preparation of Specialised Dictionaries*. Amsterdam, Philadelphia: John Benjamins.

Bergenholtz, H., & Tarp, S. (2003). Two Opposing Theories: On H. E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes*, 31, 171-196.

Bernardini, S., & Ferraresi, A. (2013). Old needs, new solutions. Comparable corpora for language professionals. In S. Sharoff, R. Rapp, P. Zweigenbaum & P. Fung (Eds.), *BUCC: Building and Using Comparable Corpora* (303-319). Berlin, Heidelberg: Springer. Retrieved from

https://www.researchgate.net/publication/280064849_Old_Needs_New_Solutions_Comparable_Corpora_for_Language_Professionals

Berti, B., & Pinnavaia, L. (2012). Towards a Corpus-driven Italian-English Dictionary of Collocations. In R. Facchinetti (Ed.), *English Dictionaries as Cultural Mines* (pp. 201-222). Newcastle-Upon-Tyne: Cambridge Scholars Publishing.

Berti, B., & Pinnavaia, L. (2014). Creating a Bilingual Italian-English Dictionary of Collocations. In: A. Abel, C. Vettori, & N. Ralli (Eds.), *Proceedings of the 16th EURALEX International Congress* (pp. 515-524). Bolzano. Retrieved from http://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202014/euralex_2014_038_p_515.pdf

Bogaards, P. (1996). Dictionaries for Learners of English. *International Journal of Lexicography*, 9 (4), 277-320.

Bogaards, P., & van der Kloot, W. (2002). Verb Constructions in Learners' Dictionaries. In A. Braasch & C. Povlsen (Eds.), *Proceedings of the 10th Euralex International Congress*, (pp. 747–757). Copenhagen: CST.

Bogaards, P., & van der Kloot, W. (2002). Verb Constructions in Learners' Dictionaries. In A. Braasch & C. Povlsen (Eds.), *Proceedings of the 10th Euralex International Congress*, (pp. 747–757). Copenhagen: CST.

Boonmoh, A. (2012). E-dictionary use under the spotlight. Student's use of pocket electronic dictionaries for writing. *Lexikos*, 22, 43-68. Retrieved from http://146.232.21.2/pub/article/viewFile/997/514

Boonmoh, A., & Nesi, H. (2008). A survey of dictionary use by Thai university staff and students, with special reference to pocket electronic dictionaries. *Horizontes de Lingüística Aplicada*, 6 (2), 79-90.

Bothma, T.J.D. (2011). Filtering and adapting data and information in an online environment in response to user needs. In P. A. Fuertes-Olivera & H. Bergenholtz (Eds.*), e-Lexicography: The Internet, Digital Initiatives and Lexicography* (pp. 71-102). London, New York: Continuum.

Bothma, T.J.D. (2011): Filtering and adapting data and information in an online environment in response to user needs. In: Fuertes-Olivera P.A., Bergenholtz H. *E-Lexicography: The Internet, Digital Initiatives and Lexicography*. London/New Delhi/New York/Sydney: Bloomsbury Academic, 71-102.

Brett, D. (2017). Collocate networks in the language of crime journalism. *Studii de lingvistica*, 7, 125-144. Retrieved from http://studiidelingvistica.uoradea.ro/docs/7-2017/pdf_uri/Brett.pdf.

Busane, M. (1990). Lexicography in Central Africa: The user perspective, with special reference to Zaïre. In R. R. K. Hartmann (Ed.), *Lexicography in Africa. Progress reports from the Dictionary Research Center Workshop at Exeter* (pp. 19-35). Exeter, England: University of Exeter Press.

Carr, M. (1997). Internet Dictionaries and Lexicography. *International Journal of Lexicography*, 10 (3), 209-230.

Chen, Y. (2010). Dictionary use in EFL learning. A contrastive study of pocket electronic dictionaries and paper dictionaries. *International Journal of Lexicography*, 23 (3), 275-306.

De Schryver, G. M. (2003). Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography*, 16 (2), 143-199.

Dengler, K., Matthes, B., & Paulus W. (2014). Berufliche Tasks auf dem deutschen Arbeitsmarkt. Eine alternative Messung auf Basis einer Expertendatenbank. *FDZ-Methodenreport. Methodische Aspekte zu Arbeitsmarktdaten*, 12. Retrieved from http://doku.iab.de/fdz/reporte/2014/MR_12-14.pdf

Dictionaries. In Text Encoding Initiative (TEI) Guidelines. Retrieved from http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html

Dziemianko, A. (2012). *Noun and Verb Codes in English Monolingual Dictionaries for Foreign Learners: A Study of Usefulness in the Polish Context*. Poznań: Adam Mickiewicz University Press.

Dziemianko, A. (2012a). On the use(fulness) of paper and electronic dictionaries. In S. Granger & M. Paquot (Eds.), *Electronic lexicography* (pp. 320-341). Oxford: Oxford University Press.

Engelberg, S., & Lemnitzer, L. (2009). *Lexikographie und Wörterbuchbenutzung* (4th ed.). Tübingen: Stauffenburg.

Engelberg, S., & Storrer, A. (2016). Typologie von Internetwörterbüchern und -portalen. In A. Klosa & C. Müller-Spitzer (Eds.), *Internetlexikografie. Ein Kompendium* (pp. 31-63). Berlin, Boston: de Gruyter.

Engelberg, S., Müller-Spitzer C., & Schmidt, T. (2016). Vernetzungs- und Zugriffsstrukturen. In A. Klosa & C. Müller-Spitzer (Eds.), *Internetlexikografie. Ein Kompendium* (pp. 153-195). Berlin, Boston: de Gruyter.

Firth, J. R. (1957). *Papers in Linguistics 1934-1951*. Oxford: Oxford University Press.

Fox, G. (1987). The case for examples. In J. Sinclair (Ed.), *Looking Up. An account of the COBUILD project in lexical computing*, (pp. 137-19). London, Glasgow: Collins.

Frankenberg-Garcia, A. (2015). Dictionaries and encoding examples to support language production. *International Journal of Lexicography*, 24 (4), 490-512.

Fuertes-Olivera, P. A. (2013). E-lexicography: The continuing challenge of applying new technology to dictionary making. In H. Jackson (Ed.), *Bloomsbury Companion to Lexicography* (pp. 323-340). London: Bloomsbury.

Fuertes-Olivera, P. A., & Tarp, S. (2014). *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography.* Berlin, Boston: de Gruyter.

Geyken, A., & Lemnitzer L. (2016). Automatische Gewinnung von lexikographischen Angaben. In A. Klosa & C. Müller-Spitzer (Eds.), *Internetlexikografie. Ein Kompendium* (pp. 197-247). Berlin, Boston: de Gruyter.

Gouws, R. (2014). Article Structures: Moving from Printed to e-Dictionaries. *Lexikos*, 24, 155-177. Retrieved from http://lexikos.journals.ac.za/pub/article/view/1256

Haas, M. R. (1962). What belongs in a bilingual dictionary? F. W. Householder & S. Saporta (Eds.), *Problems in Lexicography*, (pp. 45-50). Bloomington: Indiana University.

Hanks, P. (2012). Corpus evidence and electronic lexicography. In S. Granger & M. Paquot (Eds.), *Electronic lexicography* (pp. 57-82). Oxford: Oxford University Press. Retrieved from http://www.patrickhanks.com/uploads/5/1/4/9/5149363/hanks_2012f.pdf

Hartmann, R. R. K. (2005). Onomasiological Dictionaries in 20th-Century Europe. *Lexicographica. International Annual for Lexicography*, 21, 6-19.

Hartmann, R. R. K. (2005a). Pure or hybrid? The development of mixed genres. *Facta Universitatis,* 3, 193-208. Retrieved from http://facta.junis.ni.ac.rs/lal/lal2005/lal2005-06.pdf

Hausmann, F. J. (1984). Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortverbindungen. *Praxis des neusprachlichen Unterrichts*, 31, 385-406.

Hausmann, F. J. (1989). Wörterbuchtypologie. In F. J. Hausmann, O. Reichmann, H. E. Wiegand, L. Zgusta (Eds.), *Dictionaries: An International Encyclopedia of Lexicography* (vol. 1, pp. 968-981). Berlin, New York: de Gruyter.

Hausmann, F. J. (2007). Die Kollokationen im Rahmen der Phraseologie – Systematische und historische Darstellung. *Zeitschrift Für Anglistik und Amerikanistik*, 55, 217-234. Retrieved from http://www.zaa.uni-tuebingen.de/wp-content/uploads/05-Hausmann_ffinal.pdf

Heid, U., Prinsloo, D. J., & Bothma, T. J. D. (2012). Dictionary and corpus data in a common portal: state of the art and requirements for the future. *Lexicographica*, 28, 269-291.

Heid, U., Prinsloo, D. J., & Bothma, T. J. D. (2012). Dictionary and corpus data in a common portal: state of the art and requirements for the future. *Lexicographica*, 28, 269-291.

Herbst, T., & Stein G. (1987). Dictionary-Using Skills: a plea for a new orientation in language teaching. In A. P. Cowie (Ed.), *The Dictionary and the Language Learner* (pp. 115-127). Tübingen: Niemeyer.

Hildenbrandt, V., & Klosa, A. (Eds.). (2016). Lexikographische Prozesse bei Internetwörterbüchern. *Opal – Online publizierte Arbeiten zur Linguistik,* 1. Retrieved from https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/4730/file/Hildenbrandt_Klosa_Lexikographische_Prozesse_bei_Internetw%C3%B6rterb%C3%BCchern_2016.pdf

Hüllen, W. (1994). *The world in a list of words*. Tübingen: Niemeyer.

Hüllen, W. (1999). *English Dictionaries 800–1700: The Topical Tradition*. Oxford: Clarendon.

Jackson, H. (2002). *Lexicography: An Introduction.* London: Routledge.

Janser, M. (2018). The greening of jobs in Germany. First evidence from a text mining based index and employment register data. *IAB-Discussion Paper*, 14, 1-76. Retrieved from http://doku.iab.de/discussionpapers/2018/dp1418.pdf

Kemmer, K. (2010). Onlinewörterbücher in der Wörterbuchkritik. Ein Evaluationsraster mit 39 Beurteilungskriterien. In OPAL – Online publizierte Arbeiten zur Linguistik, 2, 1-33.

Kenning, M. (2010). What are parallel and comparable corpora and how can we use them? In M. McCarthy & A. O'Keeffe (Eds.), *The Routledge Handbook of Corpus Linguistics*, (pp. 487-501). Abingdon: Routledge.

Kilgarriff, A. (2013). Using corpora as data source for dictionaries. In H. Jackson (Ed.), *The Bloomsbury Companion to Lexicography* (pp. 77-96). London: Bloomsbury.

Kipfer, B. A. (1986). Investigating an onomasiological approach to dictionary material. *Dictionaries: Journal of the Dictionary Society of North America,* 8, 55-64.

Kirkpatrick, B. (1989). User's Guides in Dictionaries. In F. J. Hausmann, O. Reichmann, H. E. Wiegand, L. Zgusta (Eds.), *Dictionaries: An International Encyclopedia of Lexicography* (vol. 1, pp. 754-761). Berlin, New York: de Gruyter.

Klosa, A. (2013). The lexicographical process (with special focus on online dictionaries). In R. H. Gouws, U. Heid, W. Schweickard & H. E. Wiegand (Eds.). *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume. Recent developments with focus on electronic and computational lexicography* (pp. 517-524). Berlin, Boston: de Gruyter.

Klosa, A., & Müller-Spitzer, C. (Eds.). (2016). *Internetlexikografie. Ein Kompendium.* Unter Mitarbeit von Martin Loder. Berlin, Boston: de Gruyter.

Klosa, A., & Rufus, G. (2015). Outer features in e-dictionaries. *Lexicographica*, 31, 142-172. Retrieved from https://core.ac.uk/download/pdf/83653976.pdf

Koplenig, A., & Müller-Spitzer, C. (2014). General issues of online dictionary use. In C. Müller-Spitzer (Ed.). *Using Online Dictionaries* (pp. 127-141). Berlin, Boston: de Gruyter.

Kühn, P. (1989). Typologie der Wörterbücher nach Benutzungsmöglichkeiten. In F. J. Hausmann, O. Reichmann, H. E. Wiegand, L. Zgusta (Eds.), *Dictionaries: An International Encyclopedia of Lexicography* (vol. 1, pp. 111-127). Berlin, New York: de Gruyter.

Kunze, C., & Lemnitzer, L. (2007). *Computerlexikographie: Eine Einführung.* Gunter Narr Verlag: Tübingen.

L' Homme, M. (2003). Capturing the lexical structure in special subject fields with verbs and verbal derivatives. A model for spacilized lexicography. In International Journal of Lexicography 16 (4), 403-421.

Laffling, J. (1992). On constructing a transfer dictionary for man and machine. *Target*, 4(1), 17–31.

Landau S. (2001). *Dictionaries. The Art and Craft of Lexicography.* Cambridge: Cambridge University Press.

Laufer B., & Melamed L. (1994). Monolingual, bilingual and 'bilingualised' dictionaries: which are more effective, for what and for whom? In W. Martin, W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg, & P. Vossen (Eds.), *Proceedings of the 6th Euralex International Congress* (565-576). Amsterdam: Euralex.

Laufer, B. & Kimmel, M. (1997). Bilingualised dictionaries: how learners really use them. *System*, 25 (3), 361-369. Retrieved from https://www.academia.edu/10246535/Bilingualised_dictionaries_how_learners_really_use_them

Laufer, B. (1992). Corpus-based versus lexicographer examples in comprehension and production of new words. In H. Tommola, K. Varantola, T. Salmi-Tolonen, & J. Schopp (Eds.), *Euralex '92 Proceedings* (pp. 71-76). Retrieved from http://www.euralex.org/elx_proceedings/Euralex1992_1/013_Batia%20Laufer%20-Corpus-based%20versus%20lexicographer%20examples%20in%20comprehension%20and%20production%20of%20n.pdf

Laufer, B., & Waldman, T. (2011). Verb-noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning*, 61, 647-672.

Lehr, A. (1996). Zur neuen Lexicographica-Rubric "Electronic dictionaries". Lexicographica, 12, 310-317.

Lew, R. (2004). *Which dictionary for whom? Receptive use of bilingual, monolingual and semi-bilingual dictionaries by Polish learners of English*. Poznań: Motivex.

Lew, R. (2011). Online Dictionaries of English. In P. A. Fuertes-Olivera & H. Bergenholtz (Eds.*), e-Lexicography: The Internet, Digital Initiatives and Lexicography* (pp. 230-250). London, New York: Continuum.

Lew, R. (2011a). User studies: Opportunities and limitations. In A. Akasu & S. Uchida (Eds.), *Lexicography: Theoretical and practical perspectives. Papers submitted to the seventh ASIALEX Biennial International Conference* (pp.7-16). Kyoto: The Asian Association for Lexicography.

Lew, R. (2012). How can we make electronic dictionaries more effective? In S. Granger & M. Paquot (Eds.), *Electronic lexicography* (pp. 343-361). Oxford: Oxford University Press.

Lew, R. (2013). Space restrictions in paper and electronic dictionaries and their implications for the design of production dictionaries. Retrieved from https://repozytorium.amu.edu.pl/bitstream/10593/799/1/Lew_space_restrictions_in_paper_and_electronic_dictionaries.pdf

Lew, R. (2013a). Online dictionary skills. In K. Iztok, J. Kallas, P. Gantar, S. Krek, M. Langemets & M. Tuulik (Eds.), *Electronic lexicography in the 21st century: Thinking outside the paper. Proceedings of the eLex 2013 conference* (pp. 16-31). Tallinn, Estonia: Trojina, Institute for Applied Slovene Studies. Retrieved from http://eki.ee/elex2013/proceedings/eLex2013_02_Lew.pdf

Lew, R. (2014). User-generated content (UGC) in English online dictionaries. *Opal – Online publizierte Arbeiten zur Linguistik*, 4, 8-26. Retrieved from https://pub.ids-mannheim.de//laufend/opal/pdf/opal2014-4.pdf

Lew, R. (2016). Can a dictionary help you write better? A user study of an active bilingual dictionary for Polish learners of English. *International Journal of Lexicography*, 29 (3), 353–366.

Losey-León, M. A. (2015). Corpus-based Contrastive Analysis of Keywords and Collocations across Sister Specialized Subcorpora in the Maritime Transport Field*. Procedia - Social and Behavioral Sciences.* 198, 526-534.

Maingay, S., and Rundell, M. (1990). What makes a good dictionary example? *Paper presented at the 24th LATEFL Conference*. Dublin.

Maingay, S., and Rundell, M. (1990). What makes a good dictionary example? *Paper presented at the 24th LATEFL Conference*. Dublin.

McArthur, T. (1986). *Worlds of Reference*. *Lexicography, Learning and Language from the Clay Tablet to the Computer.* Cambridge: Cambridge University Press.

McArthur, T. (1998). *Living Words: Language, Lexicography and the Knowledge Revolution*. Exeter: Exeter University Press.

Meliss, M., & Sánchez Hernández, P. (2015) Theoretical and methodological foundations of the DICONALE project: a conceptual dictionary of German and Spanish. In *Planning non-existent dictionaries* (pp. 163-179). Lisboa: Centro de Linguística da Universidade de Lisboa.

Meyer, P. (2014). Meta-computerlexikografische Bemerkungen zu Vernetzung in XML-basierten Onlinewörterbüchern – am Beispiel von ELEXIKO. *Opal – Online publizierte Arbeiten zur Linguistik*, 2, 9-21. Retrieved from https://pub.idsmannheim.de/laufend/opal/pdf/opal2014-2.pdf

Mugdan, J. (1992). Zur Typologie zweisprachiger Wörterbücher. In G. Meder, & A. Därner (Eds.), *Worte, Wörter, Wörterbücher: Lexikographische Beiträge zum Essener Linguistischen Kolloquium* (pp. 25-48). Tübingen: Niemeyer.

Müller-Spitzer, C. (2003). Ordnende Betrachtungen zu elektronischen Wörterbüchern und lexikographischen Prozessen. *Lexicographica*, 19, 140-168.

Müller-Spitzer, C. (2007). *Der lexikographische Prozess. Konzeption für die Modellierung der Datenbasis.* Tübingen: Narr.

Müller-Spitzer, C. (2014). Empirical data on contexts of dictionary use. In C. Müller-Spitzer, (Ed.), *Using Online Dictionaries* (pp.85-126). Berlin, Boston: de Gruyter.

Müller-Spitzer, C. (2016). Nutzerbeteiligung. In A. Klosa & C. Müller-Spitzer (Eds.), *Internetlexikografie. Ein Kompendium* (pp. 291-342). Berlin, Boston: de Gruyter.

Müller-Spitzer, C., & Koplenig, A. (2014). Online dictionaries: expectations and demands. In C. Müller-Spitzer (Ed.). *Using Online Dictionaries* (pp. 143-188). Berlin, Boston: de Gruyter.

Nesi, H. (1999). The specification of dictionary reference skills in higher education. In R. R. K. Hartmann (Ed.), *Dictionaries in language learning. Recommendations, national reports and thematic reports from the Thematic Network Project in the Area of Languages, sub-project 9: dictionaries* (pp. 53-67). Berlin: Freie Universität Berlin. Retrieved from https://www.researchgate.net/publication/272167806_The_Specification_of_Dictionary_Reference_Skills_in_Higher_Education

Nesi, H. (2000). Electronic Dictionaries in Second Language Vocabulary Comprehension and Acquisition: the State of the Art. In: U. Heid, S. Evert, E. Lehmann & C. Rohrer (Eds.), *Proceedings of the 9th Euralex International Congress* (pp. 839-847). Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. Retrieved from http://www.euralex.org/elx_proceedings/Euralex2000/099_Hilary%20NESI_Electronic%20Dictionaries%20in%20Second%20Language%20Vocabulary%20Comprehension%20and%20Acquisition_the%20State%20of%20the%20Art.pdf

Nesi, H., & Meara P. (1994). Patterns of misrepresentation in the productive use of EFL dictionary definitions. *System*, 22 (1), 1-15.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.

Nielsen, S., & Almind, R. (2011). In P. A. Fuertes-Olivera & H. Bergenholtz (Eds.*), e-Lexicography: The Internet, Digital Initiatives and Lexicography* (pp. 141-167). London, New York: Continuum.

Nielsen, S., & Fuertes-Olivera, P. (2013). Development in Lexicography: From Polyfunctional to Monofunctional Accounting Dictionaries. *Lexikos 23*, 323-347.

Nuccorini, S. (1994). On Dictionary Misuse. In W. Martin, W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg, & P. Vossen (Eds.), *Proceedings of the 6th Euralex International Congress* (586-597). Amsterdam: Euralex.

Osselton, N.E. (2000). Murray and his European counterparts. In L. Mugglestone (Ed), *Lexicography and the OED: Pioneers in the Untrodden Forest* (pp.59-76). Oxford: Oxford University Press.

Pearson J. (1998). *Terms in Context*. Amsterdam, Philadelphia: John Benjamins.

Reichmann, O. (1990). Das onomasiologische Wörterbuch: Ein Überblick. In F. J. Hausmann, O. Reichmann, H. E. Wiegand & L. Zgusta (Eds.), *Dictionaries: An International Encyclopedia of Lexicography* (vol. 2, pp. 1057-1067). Berlin, New York: de Gruyter.

Renouf, A. (1987). Corpus Development. In Sinclair, J. (Ed.), *Looking up. An account of the Cobuild Project in lexical computing* (pp. 1-40). London: Collins.

Rundell, M. (1999). Dictionary use in production. *International Journal of Lexicography*, 12 (1), 35–53.

Rundell, M. (2002). Introduction. In *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan Education.

Rundell, M. (2010). What future of the learner's dictionaries? In I. Kernerman, & P. Bogaards (Eds.), *English Learner's Dictionaries at the DSNA 2009*, (pp. 169-175). Jerusalem: Kdictionaries.

Rundell, M. (2012). 'It works in practice but will it work in theory?' The uneasy relationship between lexicography and matters theoretical. In R. Vatvedt Fjeld & J. M. Torjusen (Eds.). *Proceedings of*

the 15th EURALEX International Congress, (pp. 47-92). Oslo: UiO. Retrieved from http://www.euralex.org/elx_proceedings/Euralex2012/pp47-92%20RundellI.pdf

Sánchez Hernández, P. (2013). Zur Konzipierung eines deutsch-spanischen kombiniert onomasiologisch-semasiologisch ausgerichteten Verbwörterbuchs mit online-Zugriff—ausgewählte Aspekte. *Aussiger Beiträge*, 7, 135–155.

Schierholz, S. J. (2015). Methods in Lexicography and Dictionary Research. *Lexicos*, 25, 323-352.

Sharoff, S., Rapp, R., Zweigenbaum, P., & Fung, P. (Eds.). (2013). *Building and using comparable corpora*. Heidelberg: Springer.

Shcherba, L.V. (1940). Towards a general theory of lexicography. *International Journal of Lexicography,* 8 (4), 1995, 315-350.

Sierra, G. & Hernández, L. (2013). Automatic construction of the knowledge base of an onomasiological dictionary. Retrieved from https://www.researchgate.net/publication/283773324_Automatic_construction_of_the_knowledge_base_of_an_onomasiological_dictionary

Sierra, G. & McNaught J. (2000) Extracting Semantic Clusters from MRDs for an Onomasiological Search Dictionary. *International Journal of Lexicography*, 13 (4), 264-286.

Sierra, G. (2000). The onomasiological dictionary: a gap in lexicography. In U.Heid et al. (Eds.). *Proceedings of the 10th Euralex International Congress*, (pp. 223-235). Stuttgart: Universität Stuttgart. Retrieved from https://www.researchgate.net/publication/265099932_The_onomasiological_dictionary_a_gap_in_lexicography

Sierra, G. (2008). Natural Language Searching in Onomasiological Dictionaries. Retrieved from https://www.researchgate.net/publication/238529713_Natural_Language_Searching_in_Onomasiological_Dictionaries.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. (1996). Preliminary Recommendations on Corpus Typology. EAGLES Document EAG-TCWG-CTYP/P. Retrieved from http://www.ilc.cnr.it/EAGLES96/corpustyp/corpustyp.html

Sinclair, J. (2004). Intuition and annotation - the discussion continues. In K. Aijmer, & B. Altenberg (Eds.). *Advances in corpus linguistics. Papers from the 23rd International Conference on English Language Research on Computerized Corpora (ICAME 23).* (pp. 38-59). Amsterdam, New York: Rodopi.

Skadina, I., Vasiļjevs, A., Skadiņš, R., Gaizauskas, R., Tufiş, D., & Gornostay, T. (2010). Analysis and Evaluation of Comparable Corpora for Under Resourced Areas of Machine Translation. *Proceedings of the 3rd Workshop on Building and Using Comparable Corpora* (pp. 6-14). Retrieved from https://www.researchgate.net/publication/228517099_Analysis_and_evaluation_of_comparable_corpora_for_under_resourced_areas_of_machine_translation

Stark, M. (2011). *Bilingual Thematic Dictionaries*. Berlin, Boston: de Gruyter.

Storrer, A., & Freese, K. (1996). Wörterbücher im Internet. *Deutsche Sprache,* 2, 97-153.

Tarp, S. (2008*). Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Niemeyer.

Tarp, S. (2009). Reflections on lexicographic user research. *Lexicos*, 20, 450-465.

Tarp, S. (2011). Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. In P. A. Fuertes-Olivera, & H. Bergenholtz (Eds.). *e-Lexicography: The Internet, Digital Initiatives and Lexicography*, (pp. 54-70). London, New York: Continuum.

Tarp, S. (2011). Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. In P. A. Fuertes-Olivera, & H. Bergenholtz (Eds.). *e-Lexicography: The Internet, Digital Initiatives and Lexicography*, (pp. 54-70). London, New York: Continuum.

Tarp, S. (2012). Theoretical challenges in the transition from the lexicographical p-works to e-tools. In S. Granger & M. Paquot (Eds.), *Electronic Lexicography* (pp. 107-118). Oxford: Oxford University Press.

Taylor, A., & Chan, A. (1994). Pocket electronic dictionaries and their use. In W. Martin, W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg, & P. Vossen (Eds.) *Proceedings of the 6th Euralex International Congress* (598-605). Amsterdam: Euralex.

Text Encoding Initiative (TEI) Guidelines. Retrieved from https://tei-c.org/

Thier, K. (2014). Das Oxford English Dictionary und seine Nutzer. *Opal – Online publizierte Arbeiten zur Linguistik,* 4, 63-70. Retrieved from https://pub.ids-mannheim.de//laufend/opal/pdf/opal2014-4.pdf

Tomaszczyk, J. (1979). Dictionaries: users and uses. *Glottodidactica*, 12, 103-119.

Tono, Y. (2010). A critical review of the theory of lexicographical functions. *Lexicon*, 40, 1-26.

Töpel, A. (2014). Review of Research into the Use of Electronic Dictionaries. In C. Müller-Spitzer (Ed.) *Using online dictionaries* (pp. 13–54). Berlin, Boston: de Gruyter.

Trap-Jensen, L. (2010). One, Two, Many: Customization and User Profiles in Internet Dictionaries. In A. Dykstra & T. Schoonheim (Eds.). *Proceedings of the 14th Euralex International Congress* (pp.1133-1143). Ljouwert: Fryske Akademy.

Verlinde, S., & Binon, J. (2010). Monitoring Dictionary Use in the Electronic Age. In A. Dykstra & T. Schooheim (Eds.), *Proceedings of the 14th EURALEX International Congress* (pp. 1144-1151). Ljouwert: Fryske Akademy.

Verlinde, S., & Peeters, G. (2012). Data access revisited: The Interactive Language Toolbox. In S. Granger & M. Paquot (Eds.), *Electronic Lexicography* (pp. 147-162). Oxford: Oxford University Press.

Welker, H. A. (2010). *Dictionary Use: A General Survey of Empirical Studies.* Brasilia: Author's edition. Retrieved from http://www.let.unb.br/hawelker/dictionary_use_research.pdf

Wiegand, H. E. (1989). Arten von Mikrostrukturen im allgemeinen Einsprachigen Wörterbuch. In F. J. Hausmann, O. Reichmann, H. E. Wiegand & L. Zgusta (Eds.), *Dictionaries: An International Encyclopedia of Lexicography* (vol.1, pp. 462-501). Berlin, New York: de Gruyter.

Wiegand, H. E. (1989a). Der gegenwärtige Status der Lexikographie und ihr Verhältnis zu anderen Disziplinen. In F. J. Hausmann, O. Reichmann, H. E. Wiegand & L. Zgusta (Eds.), *Dictionaries: An International Encyclopedia of Lexicography* (vol. 1, pp. 246-279). Berlin, New York: de Gruyter.

Wiegand, H. E. (1998). *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie.* Vol. 1. Berlin, New York: de Gruyter.

Wiegand, H. E.(1977). Nachdenken über Wörterbücher: Aktuelle Probleme. In G. Drosdowski, H. Henne & H. E. Wiegand (Eds.), *Nachdenken über Wörterbücher* (pp. 51-102). Mannheim: Bibliographisches Institut.

Wiegand, H. E., Feinauer, I., & Gouws R. (2013). Types of Dictionary Articles in Printed Dictionaries.In R. Gouws, U. Heid, W. Schweickard & H. E. Wiegand (Eds.), *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography* (pp. 314-366). Berlin, New York: de Gruyter.

Wiegand, H. E., Beißwenger, M., Gouws, R. H., Kammerer, M., Storrer, A. & Wolski, W. (2010). Systematische Einführung. In H. E. Wiegand, M. Beißwenger, R. H. Gouws, M. Kammerer, A.

Storrer & W. Wolski (Eds.) *Wörterbuch zur Lexikographie und Wörterbuchforschung*, (pp. 1-121), vol.1, A-C. Berlin, New York: de Gruyter.

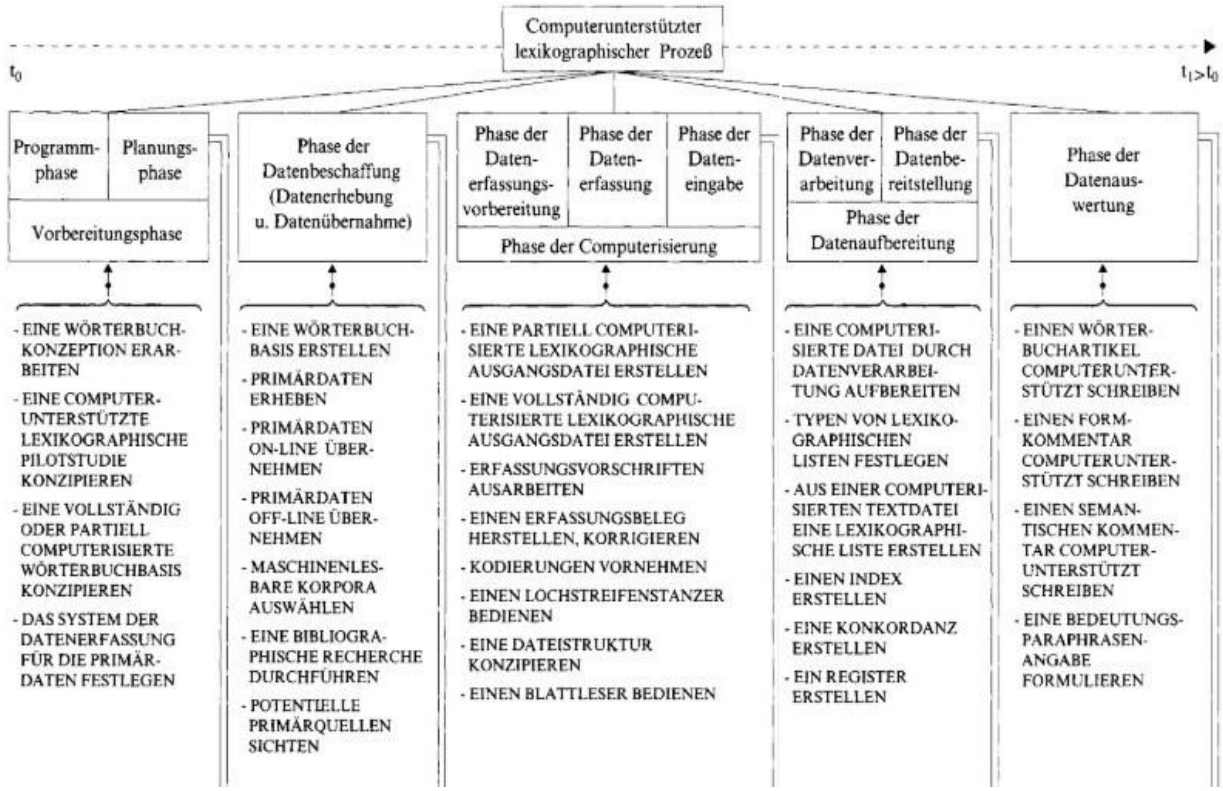
Wiegand, H.E. (2001). Was eigentlich sind Wörterbuchfunktionen? Kritische Anmerkungen zur neueren und neuesten Wörterbuchforschung. *Lexicographica,* 17, 217-248.

Wingate, U. (1999). Schwierigkeiten beim Gebrauch eines einsprachigen Lernerwörterbuchs. *Informationen Deutsch als Fremdsprache,* 26 (5), 441-457.

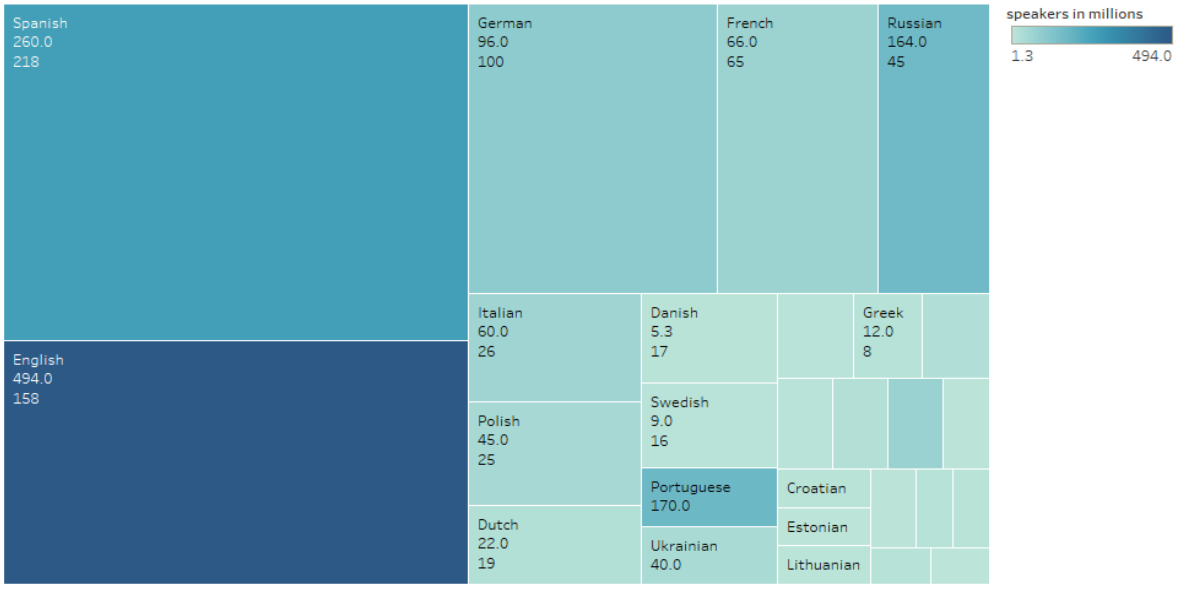
Zgusta, L. (1971). *Manual of Lexicography*. The Hague, Paris: Mouton.

Appendices

Appendix 1: Tasks for each computer-lexicographical phase (from Wiegand, 1998, p. 236)

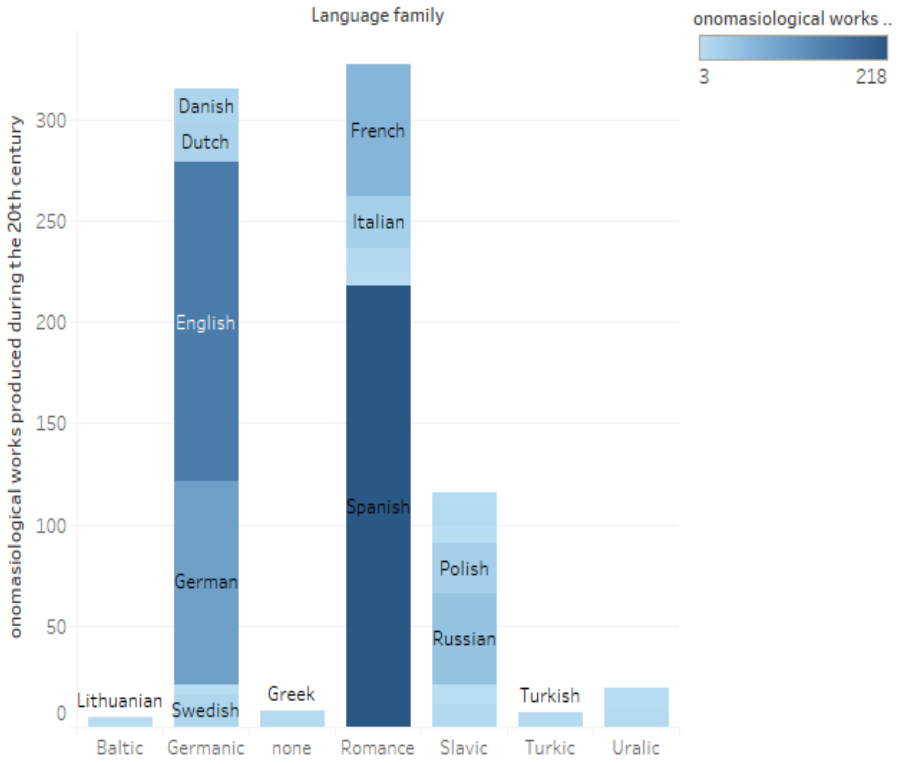


Appendix 2: Languages with the highest number of onomasiological works. Based on Hartmann (2005). Visualized in Tableau Software



Language, sum of speakers in millions and sum of onomasiological works produced during the 20th century. Color shows sum of speakers in millions. Size shows sum of onomasiological works produced during the 20th century. The marks are labeled by language, sum of speakers in millions and sum of onomasiological works produced during the 20th century. The view is filtered on language, which has multiple members selected.

Appendix 3: Number of onomasiological works per language family. Visualized in Tableau software.[1]



Sum of onomasiological works produced during the 20th century for each Language family. Color shows sum of onomasiological works produced during the 20th century. The marks are labeled by language.

[1] Hartmann (2005) does not order languages according to their families. The visualization was done in order to see which language family is the richest when it comes to onomasiological works, and which particular languages among language families have the highest number of onomasiological works.

Appendix 4: Calculation of the Mutual Information score in BYU corpora[2]

In our corpora, Mutual Information is calculated as follows:

MI = log ((AB * sizeCorpus) / (A * B * span)) / log (2)

Suppose we are calculating the MI for the collocate *color* near *purple* in BYU-BNC.

A = frequency of node word (e.g. *purple*): 1262
B = frequency of collocate (e.g. *color*): 115
AB = frequency of collocate near the node word (e.g. *color* near *purple*): 24
sizeCorpus= size of corpus (# words; in this case the BNC): 96,263,399
span = span of words (e.g. 3 to left and 3 to right of node word): 6
log (2) is literally the log10 of the number 2: .30103

MI = 11.37 = log ((24 * 96,263,399) / (1262 * 115 * 6)) / .30103

Appendix 5: Dictionary typology proposed by Kühn (1989, p. 121)

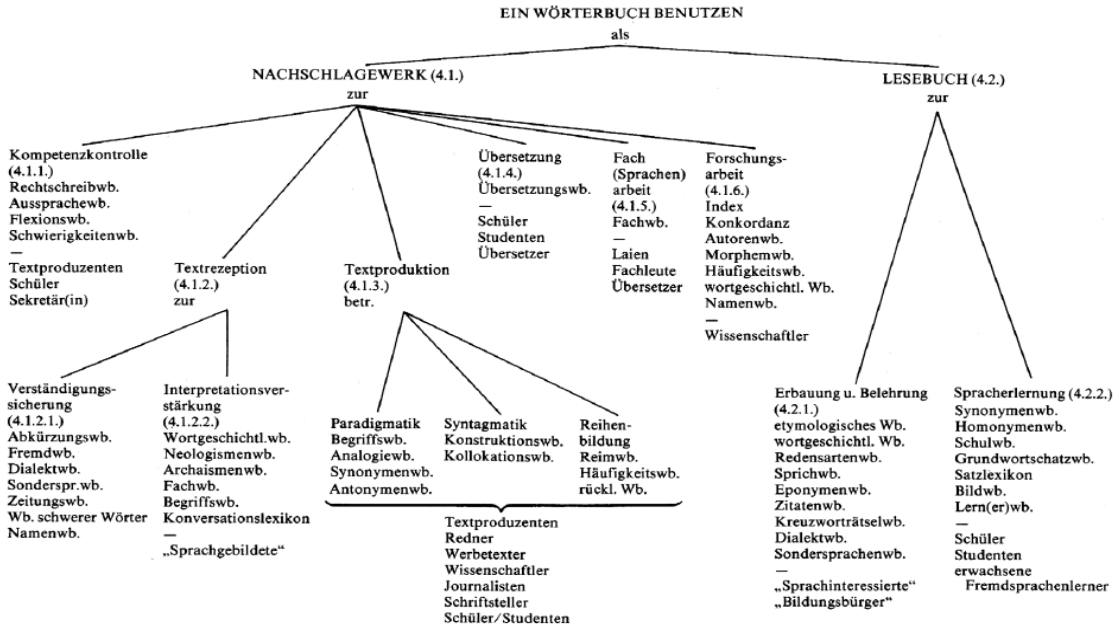


Abb. 13.1: Möglichkeiten der Wörterbuchbenutzung

[2] Retrieved from https://corpus.byu.edu/mutualInformation.asp

Appendix 6: Dictionary typology proposed by Engelberg and Lemnitzer (2009, p. 22)

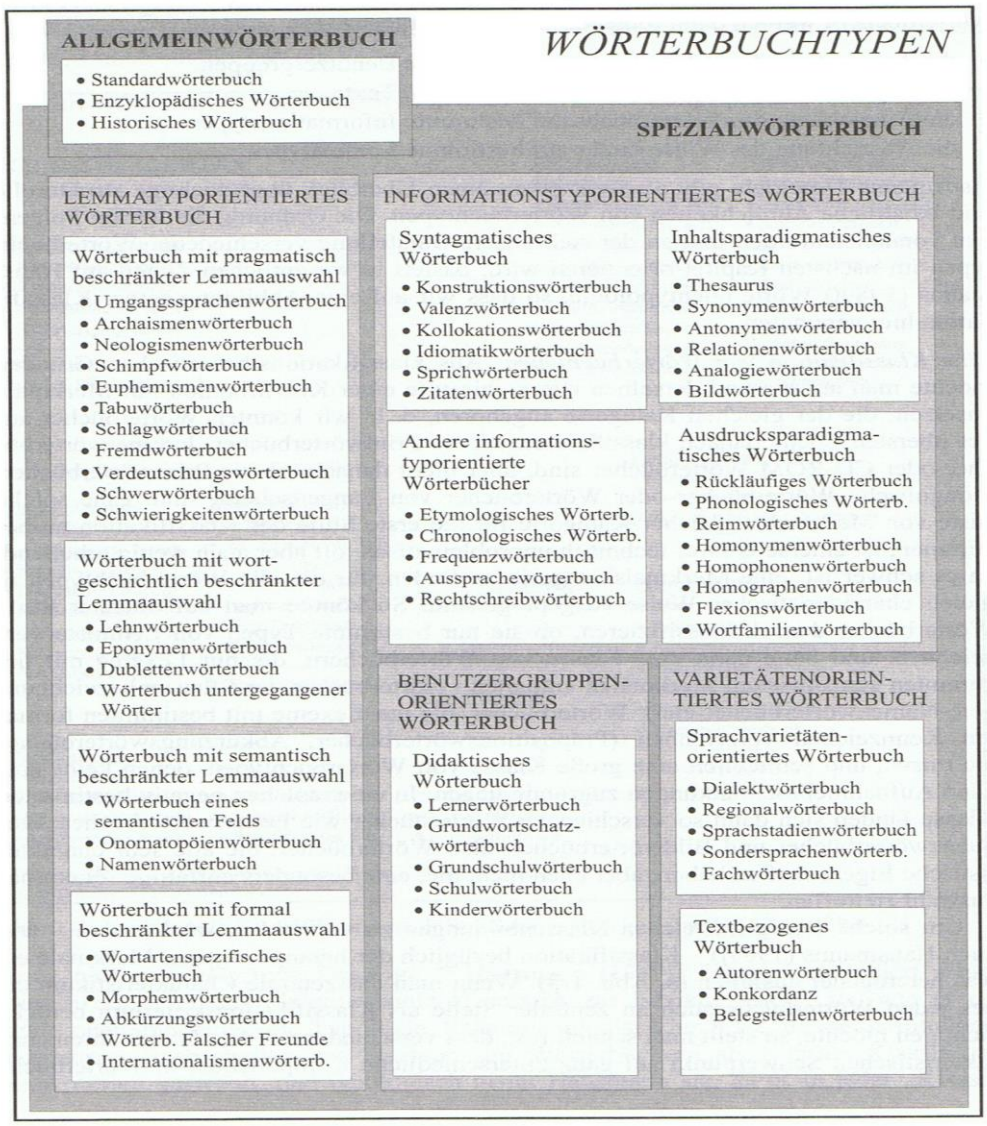
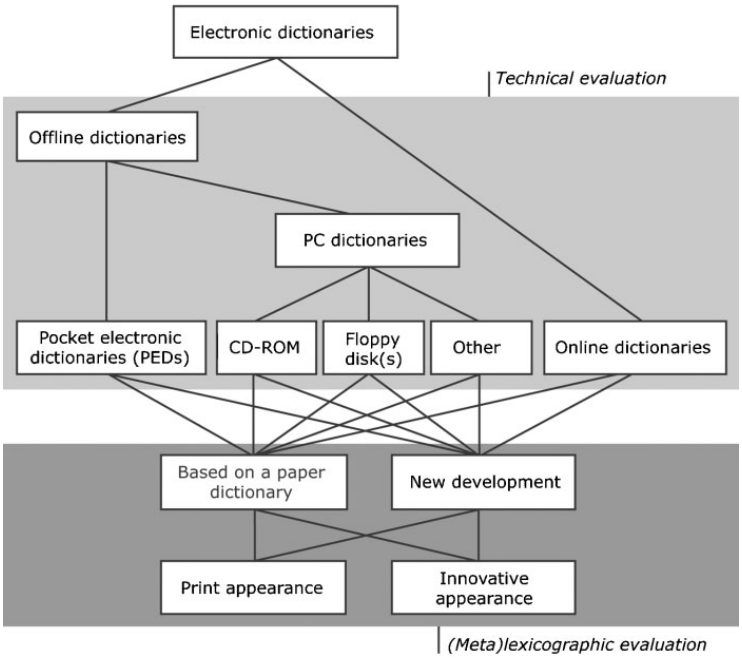
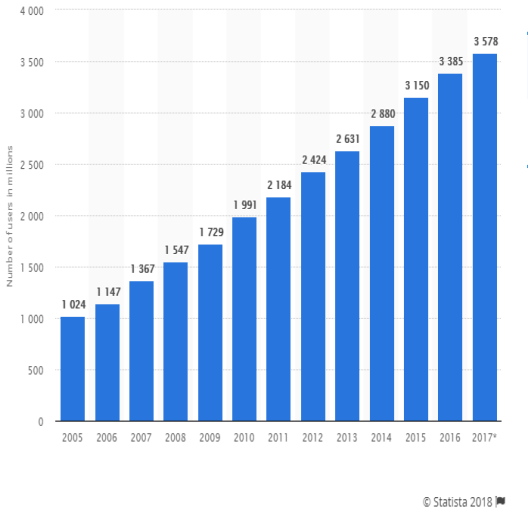


Abb. 1-3: Klassifikation von Wörterbuchtypen.

Appendix 7: Two-step technical-(meta)lexicographic electronic-dictionary typology (Lehr 1996: 315, redrawn and translated here). (de Schryver 2003, p. 148)



Appendix 8: Number of Internet users worldwide from 2005 to 2017[3]



[3] Number of internet users worldwide from 2005 to 2017 (in millions). In Statista –The portal for statistics. Retrieved from https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/

Appendix 9: Parameters for the description and evaluation of online dictionaries (Kemmer, 2010, p. 30)

	1.	**Gegenstand der Rezension**
	2.	**Adressat der Rezension**
	3.	**Perspektive des Rezensenten**
Zum OWB allgemein	4.	Titel
	5.	Wörterbuchtyp(en)
	6.	Verfasser
	7.	Herstellungsverfahren, Konzeption, Wörterbuchbasis
	8.	Abbild einer eventuellen Buchvorlage
	9.	Einbettung des OWB in (OWB-)Portal
Wörterbuchumtexte	10.	Metainformation zum OWB
	11.	Benutzungshinweise
Makrostruktur	12.	Breitenabdeckung
	13.	Makrostrukturelle Anordnung
Mikrostruktur	14.	Tiefenabdeckung
	15.	Aufbau eines Wörterbuchartikels
	16.	Grad der Textverdichtung eines Wörterbuchartikels
	17.	Multimedia
	18.	Ausdrucks- versus Begriffsinformation
Mediostruktur	19.	Interne Links
	20.	Externe Links
	21.	Konsistenz
	22.	Potenzielle Links
	23.	Formale Darstellung eines Links
	24.	Linkaktivierung
Zugriffsstrukturen & Suchmodi	25.	Retrievalmöglichkeiten
	26.	Suchfunktionen & Sucheinschränkungen
	27.	Präsentation der Suchergebnisse
	28.	Weitere Recherchemodalitäten
	29.	**Überblicksdarstellungen versus *lost in hyperspace***
Darstellung	30.	Der formale Aufbau eines Screens
	31.	Frage nach der Ästhetik – Ansprechende Gestaltung
	32.	Frage nach der Transparenz – Übersichtliche Gestaltung?
	33.	Weitere Fragen zur Gestaltung der Benutzeroberfläche
Wörterbuchbenutzung	34.	Zweck, Benutzer(gruppe[n]) & Funktion(en)
	35.	Benutzerforschung
	36.	Interaktivität I: Kommunikation Rechner-Benutzer (Datenauswahl, Lesepfade)
	37.	Interaktivität II: Kommunikation Ersteller-Benutzer (Feedback, Mitarbeit, Diskussion)
	38.	**Verfügbarkeit**
	39.	**Aktualität und Bearbeitungsstand**

Appendix 10: Questions to ask in order to identify the users' characteristics (Bergenholtz & Tarp, 2003, p. 173)

1. Which language is their mother tongue?
2. At what level do they master their mother tongue?
3. At what level do they master a foreign language?
4. How are their experience in translating between the languages in question?
5. What is the level of their general cultural and encyclopaedic knowledge?
6. At what level do they master the special-subject field in question?
7. At what level do they master the corresponding LSP in their mother tongue?

8. At what level do they master the corresponding LSP in the foreign language?

Appendix 11: Questions to ask in order to draw up lexicographically relevant user characteristics (Fuertes-Olivera & Tarp, 2014, pp. 49-50)

- *Function-relevant user characteristics:*

Which language is the user's mother tongue or the first language?

What is the user's proficiency level in the mother tongue?

With which method is the user learning the mother tongue or first language?

What is the user's proficiency level in a second, third, etc., language?

With which method is the user learning a second, third, etc., language?

What is the user's general cultural and encyclopaedic level?

What is the user's experience in translation between a specific set of languages?

What is the user's proficiency level in a specific specialized language?

What is the user's experience in translation between a specific set of specialized languages?

Etc., etc...

- *Consultation-relevant user characteristics:*

What is the user's experience of lexicographical consultations?

Is the user blind, deaf, or suffers from any other handicap which may limit the use of specific types of lexicographical tools?

Does the user have electricity and electric light?

Does the user possess a device with access to the Internet?

Does the user know how to distinguish between right and left?

Etc., etc....

Appendix 12: Six basic types of communication-oriented user situations (Bergenholtz & Tarp, 2003, p. 175)

1. Production of texts in the mother tongue (or first language)
2. Reception of texts in the mother tongue (or first language)
3. Production of texts in a foreign language (or second, third language etc.)
4. Reception of text in a foreign language (or second, third language etc.)

5. Translation of texts from the mother tongue (or first language) into a foreign language (or second, third language etc.)

6. Translation of texts from a foreign language (or second, third language etc.) into the mother tongue (or first language).

Appendix 13: Data Types in the ESSCD database[4]

Data Type	Rationale
Lemma	All dictionaries contain lemmata.
Language code to lemma	Indicates language; Assists users in identifying variants of English language
Pronunciation	Indicates the correct pronunciation of the headword
Domain and subdomain	Allows user to identify to what field belongs the headword and to access all the lemmata of a particular field or subfield.
Grammatical data addressed to lemma	Assist users in text reception and production; offers inflections, countability
Translation equivalent	Spanish equivalents assist in text reception
Language code for equivalent	Indicates the language
Grammatical data addressed to equivalent	Assist user in translation
Collocations	Assist users in production
Language code of collocation	Indicates the language
Translation of collocations	Assist users in translation and reception
Language code to translation of collocation	Indicates the language
Examples	Full sentences showing the use of collocations. Assist in production
Source	Contain bibliographical information of example sentences
Translation of examples	Full sentences showing equivalents in use. Assist in translation and reception
Notes	Assist in all user types of communicative and cognitive situations
Translation of notes	Assist in translation and reception
Images	Illustrate notes or lemmata
Cross-references	Hyperlinks to internal and external texts

[4] Based in analogy to the Table 9.2 Data Types in the Lexicographical Accounting Database (Fuertes-Olivera & Tarp, 2014, pp. 199-200).

Appendix 14: Types of summer camps[5]

Active Faith-Based Camps are operated by Christian or Jewish organizations Daily prayer, worship, and religious study are typically not major camp activities, however, may be part of daily camp life. Applicants don't have to be religious to work in these camps just need to be open-minded.

Special Needs Camps serve children and adults with disabilities. Some camps focus on providing adaptive facilities for physical limitations, some serve campers with developmental disabilities/behavioral challenges and others integrate special needs campers into traditional camp settings. Applicants don't need previous training in working with special needs, just the willingness to learn.

Underprivileged Camps are run by non-profit organizations, these traditional-style camps provide lower cost or free programming for underprivileged/disadvantaged youth who may never have been in wilderness before.

Wilderness camps: applicants will be sleeping and living in platform tents out in the wilderness.

No electricity in the cabin: the kitchen and office will have electricity for example to charge a camera and phone. The rest of the camp gives you the chance to go back to basics and forget about the stresses of everyday life whilst in nature.

Traditional Camp: are most similar to what we have seen in the movies. These camps blend a variety of sports, wilderness, creative arts and specialist activities into their daily programs.

[5] Source: Application form of Camp Canada. Retrieved from https://www.campcanada.co.uk/

Appendix 15: Collocation data for importation to Gephi as nodes table

id	label	PoS	freq
camp_NN	camp	NN	15326
summer_NN	summer	NN	4008
counsellor_NN	counsellor	NN	1606
staf f_NN	staf f	NN	3818
experience_NN	experience	NN	2819
skill_NN	skill	NN	481
development_NN	development	NN	1653
program_NN	program	NN	1588
employment_NN	employment	NN	722
member_NN	member	NN	638
interview_NN	interview	NN	614
part ċipant_NN	part ċipant	NN	1174
canoe_NN	canoe	NN	426
trip_NN	trip	NN	692
problem_NN	problem	NN	1046
solving_NN	solving	NN	386
youth_NN	youth	NN	1027
director_NN	director	NN	420
community_NN	community	NN	1739
dining_NN	dining	NN	101
hall_NN	hall	NN	113
season_NN	season	NN	265
environment_NN	environment	NN	661
life_NN	life	NN	1340
leadership_NN	leadership	NN	864
placement_NN	placement	NN	207
day_NN	day	NN	2379
team_NN	team	NN	777
session_NN	session	NN	379
training_NN	training	NN	653
applicat ȯn_NN	applicat ȯn	NN	442
process_NN	process	NN	723
sleeping_NN	sleeping	NN	105
bag_NN	bag	NN	181
age_NN	age	NN	709
group_NN	group	NN	2296
cabin_NN	cabin	NN	547
camping_NN	camping	NN	557
health_NN	health	NN	569
care_NN	care	NN	877
child_NN	child	NN	569
insurance_NN	insurance	NN	234
tripping_NN	tripping	NN	165
head_NN	head	NN	228

Appendix 16: Collocation data for importation to Gephi as edges table

source	target	type	freq
summer_NN	camp_NN	direct	1270
camp_NN	counsellor_NN	direct	993
camp_NN	staff_NN	direct	457
camp_NN	experience_NN	direct	438
skill_NN	development_NN	direct	334
camp_NN	program_NN	direct	355
camp_NN	employment_NN	direct	260
staff_NN	member_NN	direct	254
interview_NN	participant_NN	direct	166
camp_NN	participant_NN	direct	142
employment_NN	experience_NN	direct	142
canoe_NN	trip_NN	direct	136
problem_NN	solving_NN	direct	134
youth_NN	development_NN	direct	127
camp_NN	director_NN	direct	98
camp_NN	community_NN	direct	93
dining_NN	hall_NN	direct	81
camp_NN	season_NN	direct	73
camp_NN	environment_NN	direct	71
camp_NN	life_NN	direct	59
leadership_NN	skill_NN	direct	58
camp_NN	placement_NN	direct	55
day_NN	camp_NN	direct	53
staff_NN	team_NN	direct	48
camp_NN	session_NN	direct	47
staff_NN	training_NN	direct	46
leadership_NN	development_NN	direct	46
application_NN	process_NN	direct	45
sleeping_NN	bag_NN	direct	44
age_NN	group_NN	direct	42
cabin_NN	group_NN	direct	41
camping_NN	experience_NN	direct	39
health_NN	care_NN	direct	59
child_NN	care_NN	direct	59
health_NN	insurance_NN	direct	37
summer_NN	staff_NN	direct	36
child_NN	development_NN	direct	36
canoe_NN	tripping_NN	direct	35
head_NN	staff_NN	direct	35