Discursive Constructions of Culture: Semantic Modelling for Historical Travel Guides

Ulrike Czeitschner*, Barbara Krautgartner

Austrian Centre for Digital Humanities, Austrian Academy of Sciences, Austria

Copyright©2017 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Besides introducing the *Baedeker Corpus*, a digital collection of early German travel guides on non-European countries, key topics addressed in this paper are linguistic and in particular semantic annotation, domain-specific taxonomy building, corpus enrichment by embedding of external resources, and some good reasons why advanced textual studies on this genre are of interest. Early Baedeker handbooks are valuable rarities today because only a small number of copies escaped from frequent maltreatment of being cut up to save luggage weight on local trips. Those having endured give us a vivid impression of cultural narratives from the turn of the 19th century, and they tell more than one story. The presented project approaches this complex issue focusing on data development in general and the leading actors in the travel guides, *people* and *notable sights* in particular. Considering widely recognized standards for scholarly editing and relying on pan-European infrastructures to enhance data interchange and reusability, the aim is to make a well-equipped and freely accessible language resource available, which is meant to foster cross-disciplinary research on cultural representation and identity constructing discourses.

Keywords Digital Humanities, Historical Travel Guides, Digital Edition, TEI, Semantic Technologies, SKOS

1. Introduction

The *Baedeker Corpus*, which is currently being developed in the travel!digital-project¹ brings attention to historical travel guides, a part of literature often neglected by analog as well as digital humanities. In response, corpus building and

design principles aim at reducing the shortage of appropriate digital data, which allow for historical analyses, and at providing an incentive for further investigations in this field. Considering the wide distribution of travel guides in general and the cultural significance of Baedeker handbooks in particular, it is surprising that scarcely any research has yet been devoted to this sources. Outstanding quality, incontestable reliability and a very attractive appearance (see Fig. 1) contributed to their good reputation.

Karl Baedeker (1801-1859) began his publishing business in Koblenz in 1827. Appearing first in German and French, from 1861 onwards also in English, the editions soon were well known and highly influential. The *Baedeker* set the standard and defeated all competition both in- and outside the German-speaking countries. Thus, a history of travel guides without *Baedeker* would not be complete. Baedeker's first travel book, which soon became his most popular title, was a reprint of *Rheinreise von Mainz bis Cöln. Handbuch für Schnellreisende* [1], written by the German historian Johann August Klein. While this edition was extended with a course map of the river Rhine but remained unchanged otherwise, in the course of other revised editions Baedeker gave practical advices, added useful information and descriptions of the most important sights. The amendments for this volume as well as the extensive information material for other following guidebooks were based on Karl Baedeker's own travels. [cf. 2: 23 f.] Although he adopted many design elements (e.g. the red cover, the name "handbook", the asterisks) from his English contemporary John Murray, Baedeker attained the leading position in the travel book sector. At the end of the 19th century, Murray's handbooks disappeared from the European market. When business increased the regularly revised volumes were supervised by scholars and reached a high level of standardization from outer covers to inner organization. [cf. 3: 65 ff.] Numerous authors, among them historians, geographers, scientists, archaeologists, anthropologists, and linguists, guaranteed unsurpassed quality. After Karl Baedeker's death in 1859, his sons and grandsons ran the business, which still exists to this day as part of the MAIRDUMONT publishing house.

1 The project "travel!digital. Exploring *People* and *Monuments* in Baedeker Guidebooks (1875–1914)" is conducted at the *Austrian Centre for Digital Humanities ACDH*, a newly founded institute at the *Austrian Academy of Sciences*. The project is funded by the *Austrian Academy of Sciences* and the *Austrian Federal Ministry of Science, Research and Economy* within the framework of the platform *Digital Humanities Austria*.

Figure 1. Baedeker travel guides 1875-1914

Despite this longstanding success story, substantial contributions to the history of the genre² as well as its specific language of expression are still limited. In the field of textual scholarship, travel guides have not been regarded as a literary form in their own right within the classical canon for a long time. Therefore, they played a minor role in comparison to travel narratives. Complaining that “there is no general history of this significant genre of tourist literature”, Koshar [5: 15 f.] refers to the “bad reputation” travel guides have. In addition, he argues that “the variability and constantly increasing number of guidebooks frustrate the researcher’s attempt to grasp the genre as a conceptual whole”. Previous guidebook studies frequently focused on the classification of structural and functional features [cf. 6-9] and the majority of diachronic approaches is limited to one region or country.³ Since investigations rely on small analog corpora, textual and linguistic analyses as well as cultural-historical interpretations remain too restricted and usually anecdotic in nature. In contrast, the few corpus-based and corpus-driven approaches in the realm of computer linguistics that work with large data volumes, concentrate on contemporary travel guides and related texts. Often constrained to translation studies, they

exclude historical development and change [cf. 13-16]. One reason for this lack of diachronic corpus studies can be seen in the absence of appropriate electronic data, which probably is due to numerous difficulties in digitizing complexly structured historical guidebooks.

Moving away from regional restrictions and conventional methodologies, the travel!digital-project aims for a larger geographical diversification and the application of up-to-date text technology in order to create a sound basis for more systematic studies and to make this valuable part of cultural heritage available to various disciplines. Travel guides offer detailed descriptions of different places, people/s and cultures. In doing so, they do not represent neutral assessments. On the contrary, they are part of cross-cultural communication including curiosity, openness and empathy on the one hand and misunderstanding, prejudice and antagonistic relations of power on the other. Far more than mere historical sources, guidebooks mediate a variety of complex discourses. Introducing cultures and people/s, recommending itineraries, assessing sites and festivals, cultural and social conditions, suggesting modes and attitudes of behavior to adopt in these places and on these occasions at the same time disclose the authors’ own expectations and fears. In addition, they contain many interesting details illustrating what was considered age-old, traditional or modern at that time and selecting what was worth seeing.

² A notable exception is the publication of Müller [4] released in 2012.

³ Gorsemann: Iceland [6]; Pretzel [7]; Bock [10]: Rhineland; Türschau 16: Poland [11]; Epelde: India [12].

As complex inter-texts and significant discourse-historical artifacts⁴, travel guides represent “codified and authorized versions of local culture and history”. [18: 94] Reflecting dominant discourses, (re)producing and (re)constructing them, they play a central role in shaping the tourist experience and directing the tourist gaze. [cf. 19, 20] In addition to external influences, also inner qualities contribute to the complexity of the text type,⁵ which shows features of the travelogue, atlas, geographical survey, art-history guide, restaurant and hotel guide, tourist brochure, address book, and civic primer. [cf. 21: 307] For all these reasons, the various readings of history, tradition and culture including tourism and cultural heritage as well as colonialism and orientalism are of particular interest.

2. Materials and Methods

The digital collection presented in this paper brings together *all* first editions of German travel guides on non-European countries which were released by the Baedeker publishing house before World War I. The seven volumes contained, dating from the period between 1875 and 1914, comprise 4.237 pages, more than 1.7 million running words and cover various regions, offering a balanced picture of different cultural areas. (See Table 1) It was Fritz Baedeker, third son of Karl and company manager from 1869 onwards, who expanded the range to distant regions. In part, this selection reflects the project leader’s academic background in Cultural Anthropology. In part, it is due to the fact that a topical focus results in a manageable corpus size. In addition, early handbooks on non-European destinations have been treated as a minor matter in critical literature so far.

Corpus building and digitization processes (image creation, automatic text recognition, basic annotation) were already carried out in the early 2.000s within the framework of the *Austrian Academy Corpus AAC*, a predecessor institution of the *Austrian Centre for Digital Humanities*. As in the meantime structural annotation has been completed for the entire *Baedeker Corpus* as well, substantial data enrichment by means of linguistic and in particular semantic markup is at the core of the ongoing project. Aside from other well-established standards to ensure long-term preservation and free accessibility of the digitized material, we were especially interested in using the potential of semantic technologies for exploring the German

repertoire of cultural discourses at the turn of the 19th century. The project is funded by a support program of the *Austrian Academy of Sciences* that promotes sustainable digital research and data creation across humanities disciplines in accordance with international standard recommendations and best practices. Since not all of the components involved may be familiar to readers outside the DH community, the following sections giving an overview on data formats and tools, the layers of annotation, and the web application, also include introductory information on these aspects.

Table 1. *Baedeker Corpus*

1875	Palaestina und Syrien.
1877	Aegypten. Erster Theil. Unter-Aegypten bis zum Fayûm und die Sinai-Halbinsel
1891	Ägypten. Zweiter Theil. Ober-Ägypten und Nubien bis zum zweiten Katarakt
1893	Nordamerika. Die Vereinigten Staaten nebst einem Ausflug nach Mexico
1905	Konstantinopel und das westliche Kleinasien.
1909	Das Mittelmeer. Hafenplätze und Seewege nebst Madeira, den Kanarischen Inseln, der Küste Marokkos, Algerien und Tunesien
1914	Indien. Ceylon, Vorderindien, Birma, die Malayische Halbinsel, Siam, Java

2.1. Formats, Tools and Layers of Annotation

High-quality digital images produced with a *Zeutschel-OMNISCAN-7000* book scanner (8-bit, 256 greyscale) and stored in TIFF-format (400 dpi) were used as the basis for *optical character recognition OCR*, which was carried out using the software *ABBYY FineReader* [22]. As expected, changing typefaces, typographical properties and very small font sizes had a negative impact on the accuracy level achieved. Thus, text improvement was an inevitable step in each project stage. Luckily, at least semi-automatic opportunities to support time-consuming proofreading and manual corrections were available in all phases. The machine-readable text obtained from automatic recognition was exported to XML (*Extensible Markup Language*) [23], currently the state-of-the-art format for representing and sharing structured information on the Web. XML and related specifications for transforming and navigating XML documents such as XSLT (*Extensible Stylesheet Language for Transformations*) [24] and XPATH (*XML Path Language*) [25] are recommendations of the *World Wide Web Consortium W3C* [26], which is the most important international standards organization in this context.

The applied markup added to the plain text is based on the guidelines of the *Text Encoding Initiative TEI* (version P5) [27]. This standard set of rules for encoding texts of any

4 Maingueneau [17: 437] explicitly mentions travel guides as examples of discursive genres „das heißt, soziohistorisch variierende Kommunikationsdispositive“.

5 Koshar [5: 41] describes the “complexity and textual density” of Baedeker handbooks as “an intricate web of sentences and part-sentences, names of important sites in either italics or boldface (the latter with one or two asterisks if appropriate), smaller print for less essential background, references to maps and ground plans, vital dates, and historical figures. All tourist destinations found their place in a complex hierarchy of description, annotation, and meaning.”

genre defines markup elements for significant particularities at several levels and at different granularities including structural information (titles, paragraphs, verse lines, chapters, references etc.), typographical properties (highlighting, quotations etc.), and semantically distinct units (dates, named entities like persons, places etc.). Expressed in XML-syntax, TEI supports searching, text analysis and online publication as it allows to transform, rearrange, and convert the encoded texts for various presentation and navigation purposes. The entire TEI encoding scheme is extensive but at the same time modular and flexible, thus providing possibilities for customization and modification. In our case, customization simply meant that we selected a manageable subset of TEI components and generated a schema appropriate to our individual project requirements. In this context, one of our central aims was to preserve the original state of the texts in the digital representation to as high a degree as possible. Therefore, aside from structural, typographic and orthographic features, we e.g. recorded apparent errors and provided corrections for the sake of retrieval. On the phrase-level we encoded dates as well as names of persons indicating whether it is a historical, religious/mythological or literary figure. In addition, a wide range of group designations and selected sights were marked up with appropriate elements forming the groundwork for the more structured thesaurus discussed in detail below.

Both XML and TEI are application-, platform- and vendor-independent open source formats, thus there are numerous standard-compliant tools available for creating, managing, delivering, and displaying XML/TEI data and many of them are free of charge. The overall project did not develop or introduce new standards or tools but much rather used and combined existing and widely recognized components for creating a new textual resource, which is the actual result thereof.

In order to enhance comprehensive searchability, we added fundamental layers of linguistic annotation, segmenting the text into grammatical units called “tokens” (tokenization), reducing inflected word forms to a common base form called “lemma” (lemmatization), and identifying the word class of each token (part-of-speech-tagging). For the latter, we referred to the *Stuttgart-Tübingen Tag Set STTS*, a quasi-standard for tagging German texts, which includes 54 categories and therefore offers the opportunity to make detailed distinctions. [28] We implemented this step by using the *tokenEditor* [29], a sophisticated tool, developed at the *ACDH*, which allows to import TEI-conformant XML documents and to automatically enrich tokenized input data with word class and lemma information. The tool’s main and to us most important function is the possibility for manual correction of the linguistic information. In our data, mismatches are highly frequent due to historical orthography, numerous named entities and unlimited abbreviations. Thus, the various useful features of the editor like filtering, sorting and batch assignment significantly

ease the manual post processing.

In parallel, we have been working on controlled vocabularies using the *Simple Knowledge Organization System SKOS* [30] for structuring and organizing significant semantic components. “SKOS is an area of work developing specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web.” [31] SKOS adapts principles such as equivalency, hierarchical and associative relationship for expressing semantic structures, but in contrast to traditional approaches, this standard makes embedded knowledge explicit in a formalized, machine-understandable way. SKOS uses the *Resource Description Framework RDF* [32], which is another W3C-standard that provides a model for creating statements on resources and their properties in the form of subject-predicate-object expressions, also known as RDF triples. The notions *the tribe (subject) has the name (predicate) Wedda (object)* and *the Wedda (subject) live in (predicate) Ceylon (object)* are simple examples. “Asserting an RDF triple says that *some relationship, indicated by the predicate, holds between the resources denoted by the subject and object*. This statement corresponded to an RDF triple is known as an *RDF statement*” [33]. Resources, which can range from physical objects to abstract concepts, or other entities like numbers and strings, can be subject and/or object of multiple statements, thus forming a semantic network. All of the components involved are identified by *Uniform Resource Identifiers URIs* [34], which enables information systems to retrieve entities and to interpret how they are interrelated. In addition, this allows for the combination of resources from different datasets in order to enhance semantic expressiveness. Thus, SKOS/RDF supports the publication, alignment, exchange, and reuse of machine- as well as human-readable vocabularies e.g. as *Linked Data* [35].

The subtitle of the travel!digital-project „Exploring *People* and *Monuments* in Baedeker Guidebooks” summarizes the taxonomy strategy we pursued and implemented by using the SKOS standard described above: The focus on people/s and notable sights highlights two semantic fields that can be identified as prominent elements not only of travel guides but of cultural discourse itself. Both domains are essential for talking about culture and both of them are particularly suitable for data modelling.

Humans e.g. are addressed in a variety of different forms in the guidebooks. References to classes are frequently used in making generalizations about groups and individuals as well. Group members are assumed to share the same characteristic features and may therefore represent the class as a whole. [cf. 36] Each of these generic expressions is represented by a *bdk:Descriptor*, which is defined as a subclass of *skos:Concept*. Making use of the extension SKOS-XL which is particularly suitable for representing ambiguous and non-hierarchical relations below the

concept-level, individual terms are encoded as *skosxl:prefLabel(s)* and *skosxl:altLabel(s)*. Spelling variants and translations are connected on term-level by the properties *hasVariant* / *isVariantOf* and *hasTranslation* / *isTranslationOf*. Modelled in this way, the vocabulary identified in the guidebooks forms the basis of the *bdk:ConceptScheme*. To improve structuring each entry is assigned to one of the following *skos:topConcept(s)*:

- **collective terms**
Volk, Stamm, Einwohner, Einwanderer
- **ethnic / national communities**
Engländer, Japaner, Gond, Wedda
- **(more or less) geographical oriented designations**
Europäer, Asiat, Orientale, Abendländer

- **professions including political and religious functions, economic roles, and styles of living**
Händler, Gouverneur, Schamane, Bauer, Nomade
- **religiously motivated terms and names of religious communities**
Andersgläubiger, Nichthindu, Buddhist, Schivait
- **social classes**
Kaste, Klasse, Adel, Sklave

Concepts and labels include both nouns and adjectives, indicating associations among them by means of the property *skos:related*. Table 2 lists definitions of *skos:topConcept(s)* and shows selected examples of concepts, labels and terms.

Table 2. *bdk:ConceptScheme*: group designations; highlighting: interrelated concepts; variants and translations on term-level

bdk:ConceptScheme skos:hasTopConcept	bdk:Descriptor rdfs:label	skosxl:prefLabel literalForm	skosxl:altLabel literalForm
collective collective terms for groups of people	bdk:Concept1000300 Ausländer @de bdk:Concept1000500 Bevölkerung @de bdk:Concept1001000-1 einheimisch @de	bdk:Term1000300.1-de Ausländer @de bdk:Term1000500.1-de Bevölkerung @de bdk:Term1001000-1.1-de einheimisch @de	
ethnic/national names of ethnic, national communities, citizens of kingdoms, empires etc.	bdk:Concept2012200 Tamile @de bdk:Concept2012200-1 tamilisch @de *Aeneze @de	bdk:Term22012200.1-de Tamile @de <i>hasVariant</i> bdk:Term2012200-1.1-de tamilisch @de *Aeneze @de	bdk:Term2012200.2-de Tamule @de <i>isVariantOf</i> *Aenezebeduine @de
geographic (more or less) geographical oriented designations	Asiat @de Orientale @de orientalisch @de Damascener @de	Asiat @de Orientale @de orientalisch @de Damascener @de	
profession professions including political and religious functions/titles, economic roles, and styles of living	Kaufmann @de Gouverneur @de Priester @de Bauer @de Nomade @de	Kaufmann @de Gouverneur @de Priester @de Bauer @de Nomade @de	Kaufleute @de Pfarrer @de Farmer @de
religious religiously motivated terms and names of religious communities	Moslem @de buddhistisch @de	Moslem @de <i>hasTranslation</i> Muslim @en <i>isTranslationOf</i> buddhistisch @de	Mohammedaner @de Muselman @de
social names of social classes	Adel @de Bettler @de Brahmane @de proletarisch @de	Adel @de Bettler @de Brahmane @de proletarisch @de

Figure 2. Visualization of group designations in the *Baedeker Thesaurus*

The situation is similarly varied for monuments and notable sights. Since assessments and classifications of cultural heritage objects are integral parts of cultural representation, they form a separate unit in the concept scheme. Due to their infinite number, we have selected stellar sights only, which are labelled worth seeing with asterisks (*, **) in the printed editions.⁶ The topical spheres here range from architecture⁷ and artworks⁸ to accommodations, landscapes, and breath-taking views.

The systematic recording of the extensive lexical inventory identified in the travel guides resulted in the *Baedeker Thesaurus*. To give an idea of its abundance Fig. 2 shows a visualization of group designations based on two volumes, Asia Minor and India.

2.2. Data Representation in the Web Application

All of the components presented so far form essential parts of the web application we are creating. It is based on the *corpus_shell* framework [37] using the FCS/SRU protocol that is part of the CLARIN-ERIC⁹ infrastructure, an Europe-wide initiative. [38, 39] The online edition includes the digital texts together with their facsimiles and detailed metadata and combines source oriented approaches with the applied semantic technologies. [cf. 40] It provides

query capabilities inside both the text and the linguistic annotation layers. In addition, there are some classical indexes available including word forms, lemmas, word classes, and personal names. (See Fig. 3A) Optionally, their frequency distributions can be displayed in bar diagrams.

The most important and innovative feature is the integrated *Baedeker Thesaurus*. The taxonomy gives a structured overview of the rich domain-specific vocabulary. It offers definitions, reveals relationships and adds further information using external data from the *Linked Open Data LOD* cloud. For practical reasons, all entries are linked to corresponding occurrences in the *Baedeker Corpus* in order to support efficient navigation and comfort of reading. (See Fig. 3B, 3C) Transforming names of people/s and sights in the travel guides into links to the LOD cloud, the thesaurus connects occurrences in the texts to other online resources, providing enhanced access to the *Baedeker Corpus* and additional information via the guidebooks' main actors. LOD datasets such as the *Virtual International Authority File VIAF*, the *Thesaurus for the Social Sciences GESIS*, the *Art & Architecture Thesaurus AAT*, the *UNESCO Thesaurus*, and *DBpedia* will open up fresh perspectives on the genre. Aiming at easy access and data sharing, metadata creation is based on CLARIN Metadata Components. [41] Resource descriptions adapt relevant CMDI-schemas according to the project's need, providing detailed information on historical prints, the quality of the digital documents, the extent of annotations, and the applied tools. In the spirit of open access, all data will be available to others under a Creative Commons License [42] after the project's end in November 2017. Since the project is set in the framework of Austria's DARIAH-EU [43] and CLARIN-ERIC activities, corpus data and all reusable output such as the thesaurus, documentations, and schemas represent Austrian in-kind contributions to the European infrastructures.

6 Although invented by John Murray, the asterisks became famous as Baedeker brands. Murray used two stars for inns he knew from personal experience, and one star for those recommended to him. It was Karl Baedeker, who offered a whole system to lead first-time tourists to notable sites. [cf. 5: 36 f.]

7 Subcategories architecture: building of worship, cemetery, collection, educational institution, ensemble, fortification, health/sports facility, industrial building, interior design, mausoleum, monastery, multi-purpose building, office building, palace, park, ruin, scientific institution, theatre/event building, transportation facility.

8 Subcategories artworks: painting, sculpture, art collection, other artwork.

9 CLARIN Common Language Resources and Technology Infrastructure, ERIC European Research Infrastructure Consortium.

Figure 3. Baedeker Corpus Web Application. Part A, B, C

While much still remains to be done, we decided to give insight into our work at an early stage. Thus, as of December 2016, a test version of the *Baedeker Corpus* is accessible online. [44] It includes two volumes, *Konstantinopel und Kleinasien* (1905) and *Indien* (1914) and a first part of the taxonomy with about 400 sights and more than 1.000 group designations. In the course of 2017, the year of the 190th anniversary of the Baedeker Company, the remaining volumes of the corpus will be published gradually.

3. Conclusions

Along with the basic layers of linguistic annotation, the domain-specific vocabulary and content contextualization by *Linked Open Data* are appropriate and efficient instruments for exploring cultural diversity in historical travel guides from different angles. Digitally available data, fully equipped and freely accessible to interdisciplinary research, will ease exploiting such textual resources to a maximum. Without anticipating conclusions, the semantic data model presented in this paper aims at supporting fine-grained examinations of central components that have a lasting influence on cultural perceptions of “Other” and “Self”. Thus, the intended audience for this new language resource may include literary studies, discourse analyses, lexicography, linguistics, cultural anthropology, and historical geography.¹⁰ We expect that the applied semantic technologies have a great potential to reveal much about a discourse that goes far beyond travel literature.

REFERENCES

- [1] J. A. Klein. Rheinreise von Mainz bis Köln, Handbuch für Schnellreisende, Röhlings, Koblenz, 1927.
- [2] A. W. Hinrichsen. Baedeker's Reisehandbücher 1832-1990, Ursula Hinrichsen Verlag, Bevern, [1988] 1991.
- [3] J. Buzard. The Beaten Track. European Tourism, Literature and Ways to “Culture” 1800-1918, Clarendon Press, Oxford, [1993] 1998.
- [4] S. Müller. Die Welt des Baedeker. Eine Medienkulturgeschichte des Reiseführers 1830-1945, Campus Verlag, Frankfurt/New York, 2012.
- [5] R. Koschar. German Travel Cultures, Berg, Oxford, 2000.
- [6] S. Gorseman. Bildungsgut und touristische Gebrauchsanweisung. Produktion, Aufbau und Funktion von Reiseführern, Waxmann, Münster/New York, 1995.
- [7] U. Pretzel. Die Literaturform Reiseführer im 19. und 20. Jahrhundert. Untersuchungen am Beispiel des Rheins, Peter Lang, Frankfurt am Main, 1995.
- [8] W. Ramm. Textual Variation in Travel Guides, in: Ventola, Eija (ed.): Discourse and Community. Doing Functional Linguistics, Gunter Narr Verlag, Tübingen, 147-165, 2000.
- [9] K. Mittl. Baedekers Reisehandbücher. Funktionen und Bewertungen eines Reisebegleiters des 19. Jahrhunderts, Alles Buch. Studien der Erlanger Buchwissenschaft XXII. Erlangen-Nürnberg, 2007.
- [10] B. Bock. Baedeker & Cook — Tourismus am Mittelrhein 1756 bis ca. 1914, Peter Lang GmbH, Frankfurt am Main et al, 2010.
- [11] Tüschau 16. Die Darstellung anderer Kulturen. Ermittlung von Stereotypen in deutschen Polen-Reiseführern (der Jahre 1990-1996), Athena Verlag, Oberhausen, 1998.
- [12] K. R. Epelde. Travel Guidebooks to India: A Century and a Half of Orientalism, PhD, University Wollongong, 2004. Online Available: <http://ro.uow.edu.au/cgi/viewcontent.cgi?article=1195&context=theses>
- [13] S. Neumann. Textsorten und Übersetzen. Eine Korpusanalyse englischer und deutscher Reiseführer, Peter Lang, Frankfurt am Main, 2003.
- [14] P. Y. W. Lam. A corpus-driven lexico-grammatical analysis of English tourism industry texts and the study of its pedagogic implications in English for Specific Purposes, in: Hildalgo/Quereda/Santana (eds.): Corpora in the Foreign Language Classroom, Edition Rodopi B. V., 71-88, 2007.
- [15] S. Gandin. Translating the Language of Tourism. A Corpus Based Study on the Translational Tourism English Corpus (T-TourEC), in: Procedia — Social and Behavioral Sciences, Volume 95. Elsevier Ltd., 325-335, 2013.
- [16] S. Gandin. Investigating loan words and expressions in tourism discourse: a corpus driven analysis on the BBC-Travel Corpus, in: European Scientific Journal, edition vol. 10, No. 2, 1-17, 2014.
- [17] D. Maingueneau: Diskurs und Äußerungsszene. Zur gattungsspezifischen Kontextualisierung eines Zeitungsartikels zum unternehmerischen Bildungsdiskurs, in: Angermüller/Nonhoff/Herschinger et al (eds.): Diskursforschung. Ein interdisziplinäres Handbuch (2 Bde.), transcript Verlag, Bielefeld, 433-453, 2014.
- [18] A. Jaworski and A. Pritchard (eds). Discourse, Communication and Tourism, Channel View Press, Clevedon, 2005.
- [19] A. Jaworski and C. Thurlow. Scripting global discourse: The commodification of local linguacultures in tourist guidebooks, Paper presented at the annual meeting of the International Communication Association, TBA, San Francisco, CA, May 23, 2007. Online Accessible: http://citation.allacademic.com/meta/p_mla_apa_research_citation/1/7/0/4/7/pages170478/p170478-1.php
- [20] J. Urry. The Tourist Gaze. Leisure and Travel in Contemporary Societies. Sage Publications, London, [1990] 2002.
- [21] A. Steinecke. Kulturtourismus. Marktstudien – Fallstudien – Perspektiven, Oldenburg Verlag, Wien/München, 307, 2007.

¹⁰ Literary studies: genre|text type, structure|function; discourse analyses: tourism discourse, orientalism|orientalizing through textual representation, colonial discourse, cultural heritage discourse; lexicography|linguistics: person|time|place deixis, onomastics|foreign-language equivalents, foreign term translation|assimilation.

- [22] ABBYY Fine Reader. Online Available: <http://www.abbyeu.com/>
- [23] Extensible Markup Language XML. Online Available: <https://www.w3.org/XML/>. See also <https://www.w3.org/standards/xml/>
- [24] Extensible Stylesheet Language for Transformations XSLT. Online Available: <https://www.w3.org/TR/xslt>. See also <https://www.w3.org/standards/xml/>
- [25] XML Path Language XPATH. Online Available: <https://www.w3.org/TR/xpath/>. See also <https://www.w3.org/standards/xml/>
- [26] World Wide Web Consortium W3C. Online Available: <https://www.w3.org/>
- [27] Text Encoding Initiative. TEI Guidelines for Electronic Text Encoding and Interchange, Online Available: <http://www.tei-c.org/Guidelines/>
- [28] A. Schiller, S. Teufel et al. Guidelines für das Tagging deutscher Textcorpora mit STTS, 1999. Online Available: <http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-1999.pdf>
- [29] ACDH tokenEditor. Online Available: <http://www.oeaw.ac.at/acdh/en/tokenEditor>
- [30] Simple Knowledge Organization System SKOS. Online Available: <https://www.w3.org/2004/02/skos/>
- [31] <https://www.w3.org/2004/02/skos/intro>
- [32] Resource Description Framework RDF. Online Available: <https://www.w3.org/RDF/>
- [33] <https://www.w3.org/TR/rdf11-concepts/#bib-RFC3986>
- [34] Uniform Resource Identifies URI. Online Available: <https://tools.ietf.org/html/rfc3986>
- [35] Linked Data. Online Available: <http://linkeddata.org/>
- [36] D. Schmidt-Brücken. Verallgemeinerung im Diskurs. Generische Wissensindizierung in kolonialem Sprachgebrauch, Walter de Gruyter GmbH, Berlin/München/Boston, 2015.
- [37] M. Ďurčo and K. Mörth. corpus-shell. Vienna, n.d. Online Available: https://clarin.oeaw.ac.at/corpus_shell
- [38] CLARIN ERIC. Federated Content Search (CLARIN-FCS). n.d. Online Available: <https://www.clarin.eu/content/federated-content-search-clarin-fcs>
- [39] H. Stehouwer, M. Ďurčo et al. Federated Search: Towards a Common Search Infrastructure, in: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'2012), ELRA, 3255–3259, 2012. Online Available: <http://www.lrec-conf.org/proceedings/lrec2012/index.html>
- [40] A. Meroño-Peñuela, A. Ashkpour et al. Semantic Technologies for Historical Research: A Survey, Semantic Web 2014, IOS Press Journal, 1-27. <http://www.semantic-web-journal.net/content/semantic-technologies-historical-research-survey-0>
- [41] CLARIN Metadata Components CMDI. Online Available: <https://www.clarin.eu/content/component-metadata>
- [42] Creative Commons Attribution-ShareAlike 4.0 International License (CC-BY-SA 4.0). Online Available: <http://creativecommons.org/licenses/by-sa/4.0/>
- [43] Digital Research Infrastructure for the Arts and Humanities DARIAH-EU. Online Available: <http://dariah.eu/>
- [44] Baedeker Corpus. Digital Edition 2017. Ulrike Czeitschner and Barbara Krautgartner (eds.) Online Available: <https://baedeker.acdh.oeaw.ac.at/>