

# Menschen und Monumente im Fokus. Semantische Modellierung im Baedeker Corpus

,  
ulrike.czeitschner@oeaw.ac.at  
Österreichische Akademie der Wissenschaften, ACDH

## Forschungskontext

Im Rahmen des Projekts „travel!digital“ wird erstmals der Versuch unternommen, die vielfältige Terminologie in historischen Reiseführern systematisch zu erschließen. Der Fokus auf *Menschen* und *Monumente* stellt zwei dominante semantische Felder in den Mittelpunkt, die sowohl Schlüsselemente der Textsorte Reiseführer als auch wesentliche Komponenten kultureller Narrative darstellen. Damit sind Reiseführer keineswegs bloße historische Quellen, sondern vielmehr als bedeutende diskurshistorische Artefakte zu betrachten. Zudem eignet sich das in ihnen enthaltene enzyklopädische Wissen in besonderem Maße für die Datenmodellierung. Vor diesem Hintergrund erstaunt das Fehlen systematischer diachroner Forschung auf diesem Gebiet mindestens genauso wie der eklatante Mangel an digitalen Ressourcen, die für historische Untersuchungen geeignet wären.

Mit dem *Baedeker Corpus*, einer digitalen Sammlung deutschsprachiger Reiseführer aus dem Zeitraum 1875–1914, zielt das Projekt nicht nur darauf ab, wertvolles kulturelles Erbe mit den Methoden der *Digital Humanities* nachhaltig und langfristig sicherzustellen, sondern besonders darauf, semantische Technologien verstärkt zur Erforschung des deutschsprachigen Repertoires kultureller Diskurse an der Wende vom 19. zum 20. Jahrhundert einzusetzen. Als „komplexe Intertexte“ (vgl. Wierlacher 1997; Koshar 2000), die dominante Diskurse (re)produzieren bzw. (re)konstruieren, stellen Reiseführer „kodifizierte und autorisierte Versionen lokaler Kultur und Geschichte“ dar (vgl. Jaworski / Pritchard 2005). Die in den Reiseführern transportierten zeitgenössischen Lesarten zu Tourismus und kulturellem Erbe sowie Orientalismus und kolonialem Diskurs sind kultur- und diskurshistorisch von besonderem Interesse. Darüber hinaus lassen sich strukturell, linguistisch und insbesondere semantisch erschlossene digitale Reiseführer im Rahmen vergleichender literaturwissenschaftlicher Forschung, der historischen Lexikographie und Linguistik, der historischen Geographie und Kulturanthropologie nutzen.

Das vorliegende Korpus vereint *alle* Erstauflagen zu außereuropäischen Reisezielen aus dem Hause Baedeker,

die vor der Zäsur des Ersten Weltkriegs erschienen sind. Es umfasst mehr als 1,5 Mio. *running words* und deckt eine Vielzahl an Regionen ab. Mit Blick auf differenzierte Analysemöglichkeiten stehen neben der linguistischen Basis-Annotation der Volltexte (Lemmatisierung, *Part-of-Speech-Tagging*) der Aufbau kontrollierter Vokabulare und deren Anbindung an LOD-Ressourcen im Mittelpunkt. Die semantischen Repräsentationen werden mithilfe des RDF-basierten *Simple Knowledge Organization System* SKOS und dessen Erweiterung *SKOS-XL* realisiert, mit dem sich auch ambige lexikalische Einheiten und nicht-hierarchische Relationen sinnvoll organisieren lassen. Zur Erstellung der SKOS-Repräsentation des *Baedeker Corpus* wird der *OpenSKOS* Editor eingesetzt; eine Entwicklung des *Meertens Institute*, die bereits für *CLAVAS* (*CLARIN OpenSKOS Vocabulary Service*) verwendet wird, das u. a. die *ISO-639-3 language codes* im SKOS Format enthält. Auch die *DARIAH-EU* Arbeitsgruppe *Thesaurus Maintenance* verwendet *OpenSKOS* für eine erste Version der SKOS-Repräsentation des in Entwicklung befindlichen *Backbone Thesaurus*. Der *OpenSKOS* Editor sichert daher die Kompatibilität mit aktuellen Standards und Entwicklungen.

## bdk:ConceptScheme(s)

Neben Personennamen beinhalten Reiseführer auf Seiten der *Menschen* eine Vielzahl an Gruppenbezeichnungen, die generische Referenzen und als solche, Subjekte charakterisierender Eigenschaftszuschreibungen darstellen (vgl. Schmidt-Brücken 2015). Diese für diskurshistorische Analysen relevanten Belege des Sprachgebrauchs werden jeweils von einem *bdk:Descriptor*, einer Unterklasse von *skos:Concept*, repräsentiert. Ihnen zugeordnet sind einzelne Terme als *skosxl:prefLabel* und *skosxl:altLabel*. Mithilfe der *Properties hasTranslation / isTranslationOf* und *hasVariant / isVariantOf* finden die in den Reiseführern enthaltenen Übersetzungen und Varianten auf Ebene der Terme Berücksichtigung. Auf diese Weise wird der Wortschatz vollständig erfasst und in einem *skos:ConceptScheme* zusammengefasst. Der Strukturierung dieses *bdk:ConceptSchemeGroups* dienen sechs mit *skos:topConceptOf* zugeordnete Kategorien: 1) allgemeine Sammelbegriffe, 2) ethnisch/nationale Gruppen, 3) geographische Konzepte im weitesten Sinne, 4) Berufsgruppen, wozu auch politische, religiöse und wirtschaftliche Funktionen sowie Lebensstile zählen, 5) Religionsgemeinschaften und 6) soziale Gruppierungen. Aufgenommen werden als eigene Konzepte nicht nur Nomen, sondern auch Adjektive, wobei die entsprechenden Konzepte mit *skos:related* aufeinander bezogen sind. Abbildung 1 listet die *skos:topConcept(s)* und ihre Definitionen auf und gibt einige Beispiele der ihnen zugeordneten Konzepte und Labels sowie deren Beziehungen.

bdk:ConceptSchemeGroups skos:topConceptOf	bdk:Descriptor rdfs:label	skos:prefLabel literalForm	skos:altLabel literalForm
collective generic term for a group of people	bdk:Concept/00005 Ausländer @de bdk:Concept/00017 Bevölkerung @de bdk:Concept/00038 einheimisch @de	bdk:Term/00005 Ausländer @de bdk:Term/00017 Bevölkerung @de bdk:Term/00038 einheimisch @de	
ethnicNational name of an ethnic or national community	bdk:Concept/00113 Ägypter @de ↕ bdk:Concept/00115 ägyptisch @de ... 'Aeneze @de	bdk:Term/00113 Ägypter @de hasTranslation bdk:Term/00113-en Egyptian @en isTranslationOf bdk:Term/00115 ägyptisch @de hasVariant 'Aeneze @de hasVariant	bdk:Term/00114 Ägypter @de isVariantOf bdk:Term/00116 ägyptisch @de isVariantOf 'Aenezebeduine @de isVariantOf
geographic (more or less) geographical concept	... Asiate @de Orientale @de ↕ orientalistisch @de Damascener @de	... Asiate @de Orientale @de ... orientalistisch @de Damascener @de	
profession profession, political, religious, economic role, style of living	... Kaufmann @de Gouverneur @de Priester @de Bauer @de Nomade @de	... Kaufmann @de Gouverneur @de Priester @de Bauer @de Nomade @de	... Kaufleute @de Pfläner @de Farmer @de
religious name of a religious community	... Muslim @de ... buddhistisch @de	... Muslim @de hasTranslation Muslim @en isTranslationOf buddhistisch @de	... Mohammedaner @de Musliman @de
social name of a social group	... Adel @de Bettler @de Brahmane @de proletarisch @de	... Adel @de Bettler @de Brahmane @de proletarisch @de	

*Baedeker Group Taxonomy. Farbliche Hervorhebungen*  
*Spalte 2: Konzepte, die in Beziehung zu anderen*  
*Konzepten stehen; farbliche Hervorhebungen kursiv*  
*Spalte 3 und 4: Properties für Übersetzungen und*  
*Varianten auf Ebene der Terme.*

Ein ähnlich vielfältiges Bild ergibt sich im Bereich der Monumente und Sehenswürdigkeiten, die als Gegenstand wertender Beschreibung und Einordnung eine zentrale Rolle in Reiseführern einnehmen. Aufgrund ihrer Vielzahl werden zunächst nur jene Sehenswürdigkeiten strukturiert, die mit Baedeker-Sternen als besonders sehenswert gekennzeichnet sind. Das Spektrum reicht von Architektur und Kunst bis zu Unterkünften, Landschaften und atemberaubenden Aussichten. Diese Kategorien strukturieren jeweils als *skos:topConcept(s)* das *bdk:ConceptSchemeMonuments*. Derzeit wird an einer Lösung gearbeitet, die es erlaubt, zwischen profaner und sakraler Kunst und Architektur zu unterscheiden und, sofern sakral, die jeweilige Religion zuzuordnen.

## Ausblick

Die linguistische Annotation des *Baedeker Corpus* und die Erstellung der vorgestellten kontrollierten Vokabulare stellen die Voraussetzung für eine weitergehende systematische Analyse der Textsorte Reiseführer dar. Neben der Identifikation der in den untersuchten Baedeker-Bänden genannten historischen Persönlichkeiten mithilfe der *Virtual International Authority Files* und der *Deutschen Biographie* erscheint insbesondere die Anbindung der hier beschriebenen Taxonomien an Ressourcen wie die *DBpedia*, den *GESIS Thesaurus Sozialwissenschaften*, den *AAT Art & Architecture Thesaurus* des *Getty Research Institute* oder den *UNESCO Thesaurus* erfolversprechend, zumal die Verschränkung der lexikalischen Bestandsaufnahme mit der Kontextualisierung dieser externen Quellen neue

Perspektiven auf den Text zu eröffnen vermag. Vor allem aber kann das hier vorgestellte Datenmodell die granulare Analyse jener semantischen Komponenten unterstützen, die das Sprechen über sowohl das ‚Fremde‘ als auch das ‚Eigene‘ bestimmen, und erschließt somit mit rezenten Technologien die Funktionsweise eines Diskurses, dessen Wirkungsgrad bis heute weit über das Feld der Reiseliteratur hinausreicht.

Die digitalen Texte inklusive der Faksimiles und die SKOS-Vokabulare werden zu Projektende in Form einer Web-Applikation zur Verfügung gestellt. Die Navigations- und Abfragemöglichkeiten sowohl in den Volltexten als auch der linguistischen Annotation werden ergänzt durch Register der Wortformen, der Lemmata sowie der semantischen Komponenten. Die vorgestellten SKOS-Vokabulare zu *Menschen* und *Monumenten* fungieren zudem als Verbindung der konkreten Belegstellen im *Baedeker Corpus* und externen LOD-Ressourcen.

## Fußnoten

1. Das Projekt „travel!digital. Exploring *People* and *Monuments* in Baedeker Guidebooks (1875–1914)“ wird im Rahmen der Plattform *Digital Humanities Austria* gefördert.
2. Dominique Maingueneau (2014: 437) nennt Reiseführer explizit als Vertreter sogenannter *Diskursgenres*, „[...] das heißt, soziohistorisch variierende Kommunikationsdispositive“.
3. Gründe dafür finden sich u. a. darin, dass die Textwissenschaften in der Vergangenheit stets der Reiseliteratur den Vorrang gegenüber den Gebrauchstextsorten eingeräumt haben. Koshar, der das Fehlen einer allgemeinen Geschichte des Genres beklagt, verweist auf den schlechten Ruf der Reiseführer und bringt dies u. a. damit in Zusammenhang, dass die Textsorte äußerst variabel und daher konzeptuell schwer fassbar ist (Koshar 2000: 15-16). Eine Einschätzung, die eineinhalb Jahrzehnte später noch immer zutreffend ist. Als erwähnenswerte Ausnahme sei die Arbeit von Sabine Müller (2012) genannt. Dass die digitalen Geisteswissenschaften bisher kaum zur Verbesserung der Lage beigetragen haben, hängt vermutlich damit zusammen, dass die (Retro-)Digitalisierung der komplex strukturierten historischen Bände sehr aufwändig ist. Derzeit stehen wenige exemplarische Analysen kleiner analoger Korpora mit Fokus auf entweder Textsortenmerkmalen (vgl. Gorsemann 1995; Pretzel 1995; Ramm 2000; Mittl 2007) oder der Entwicklung einzelner Regionen (Gorsemann 1995: Island; Pretzel 1995: Rheinland; Forschungsgruppe Tüschau 16 1998: Polen; Epelde 2004: Indien; Bock 2010: Rheinland), noch weniger computerlinguistischen Arbeiten gegenüber. Die letzteren basieren zwar auf umfangreicheren digitalen Datenmengen, nachdem jedoch ausschließlich rezent Material herangezogen wird, bleiben historische Aspekte

unberücksichtigt (vgl. Neumann 2003; Lam 2007; Gandin 2013, 2014).

4. Für die breite Untersuchung kultureller Narrative war es naheliegend, Reiseführer zu außereuropäischen Destinationen in das Korpus aufzunehmen. Es war Fritz Baedeker, dritter Sohn des Verlagsgründers Karl Baedeker und seit 1869 Leiter des Hauses, der das Verlagsprogramm um außereuropäische Titel erweiterte: Palästina und Syrien (1875), Unter- (1877) und Ober-Ägypten (1891), Nordamerika und Mexiko (1893), Konstantinopel und Kleinasien (1905), die afrikanische Mittelmeerküste (1909) und Indien und Ceylon (1914). Die strukturelle und deskriptive XML-Annotation nach TEI-Richtlinien (Version P5) konnte bereits für das gesamte Korpus abgeschlossen werden.

5. Aktuell befindet sich der *OpenSKOS* Editor in Überarbeitung. Die neue um *SKOS-XL* Komponenten erweiterte Version soll mit Jahresende zur Verfügung stehen.

6. Unterschieden werden: Kapelle, Kirche, Kloster, Mausoleum, Friedhof, Bildungs- und Wissenschaftseinrichtung, Gesundheits- und Sporteinrichtung, Museum, Sammlung, Palast, Theater, Industriebau, Inneneinrichtung, Verkehrsbau, Ensemble, Park.

7. Unterschieden werden: Denkmal, Skulptur, Gemälde, anderes Kunstwerk, Sammlung.

8. Die Applikation basiert auf der *corpus\_shell* (#ur#o et al.), einem modularen Framework für die Publikation von Sprachressourcen. Zum FCS / SRU-Protokoll siehe CLARIN ERIC sowie Stehouwer et al. 2012.

## Bibliographie

**Bock, Benedikt** (2010): *Baedeker & Cook — Tourismus am Mittelrhein 1756 bis ca. 1914*. Bern / Berlin / Frankfurt am Main / New York / Paris / Wien: Peter Lang.

**CLARIN ERIC** (o. J.): *Federated Content Search (CLARIN-FCS)*. <https://www.clarin.eu/content/federated-content-search-clarin-fcs> [letzter Zugriff 10.09.2015].

**Digital Humanities Austria:** <http://clarin.arz.oeaw.ac.at/dha/>.

**#ur#o, Matej / Mörrth, Karlheinz / Schopper, Daniel / Siam, Omar** (o. J.): *corpus-shell*. [https://clarin.oeaw.ac.at/corpus\\_shell](https://clarin.oeaw.ac.at/corpus_shell) [letzter Zugriff 10. September 2015].

**Epelde, Kathleen R.** (2004): *Travel Guidebooks to India. A Century and a Half of Orientalism*. PhD, University Wollongong <http://ro.uow.edu.au/cgi/viewcontent.cgi?article=1195&context=theses> [10. September 2015].

**Forschungsgruppe Tüschau 16** (1998): *Die Darstellung anderer Kulturen*. Ermittlung von Stereotypen in deutschen Polen-Reiseführern (der Jahre 1990-1996). Oberhausen: Athena Verlag.

**Gandin, Stefania** (2013): "Translating the Language of Tourism. A Corpus Based Study on the Translational Tourism English Corpus (T-TourEC)", in: *Procedia — Social and Behavioral Sciences* 95: 325-335.

**Gandin, Stefania** (2014): "Investigating loan words and expressions in tourism discourse: a corpus driven analysis on the BBC-Travel Corpus", in: *European Scientific Journal* 10, 2: 1-17.

**Gorsemann, Sabine** (1995): *Bildungsgut und touristische Gebrauchsanweisung*. Produktion, Aufbau und Funktion von Reiseführern. Münster / New York: Waxmann.

**Jaworski, Adam / Pritchard, Anette** (eds.) (2005): *Discourse, communication and tourism*. Clevedon: Channel View Press.

**Koshar, Rudy** (2000): *German Travel Cultures*. Oxford: Berg.

**Lam, Peter Y. W.** (2007): "A corpus-driven lexicogrammatical analysis of English tourism industry texts and the study of its pedagogic implications in English for Specific Purposes", in: Hildalgo, Encarnación / Quereda, Luis / Santana, Juan (eds.): *Corpora in the Foreign Language Classroom*. Amsterdam / New York: Rodopi 71-88.

**Maingueneau, Dominique** (2014): „Diskurs und Äußerungsszene. Zur gattungsspezifischen Kontextualisierung eines Zeitungsartikels zum unternehmerischen Bildungsdiskurs“, in: Angermüller, Johannes / Nonhoff, Martin / Herschinger, Eva / Macgilchrist, Felicitas / Reisigl, Martin / Wedl, Juliette / Wrana, Daniel / Ziem, Alexander (eds.): *Diskursforschung*. Ein interdisziplinäres Handbuch. Bielefeld: transcript 433-453.

**Mittl, Katja** (2007): *Baedekers Reisehandbücher Funktionen und Bewertungen eines Reisebegleiters des 19. Jahrhunderts (= Alles Buch. Studien der Erlanger Buchwissenschaft 22)*. Friedrich-Alexander-Universität Erlangen-Nürnberg.

**Müller, Sabine** (2012): *Die Welt des Baedeker*. Eine Medienkulturgeschichte des Reiseführers 1830-1945. Frankfurt / New York: Campus Verlag.

**Neumann, Stella** (2003): *Textsorten und Übersetzen*. Eine Korpusanalyse englischer und deutscher Reiseführer. Bern / Berlin / Frankfurt am Main / New York / Paris / Wien: Peter Lang.

**Pretzel, Ulrike** (1995): *Die Literaturform Reiseführer im 19. und 20. Jahrhundert*. Untersuchungen am Beispiel des Rheins. Bern / Berlin / Frankfurt am Main / New York / Paris / Wien: Peter Lang.

**Ramm, Wiebke** (2000): "Textual Variation in Travel Guides", in: Ventola, Eija (ed.): *Discourse and Community*. Doing Functional Linguistics. Tübingen: Gunter Narr 147-165.

**Schmidt-Brücken, Daniel** (2015): *Verallgemeinerung im Diskurs*. Generische Wissensindizierung in kolonialem Sprachgebrauch. Berlin / München / Boston: Walter de Gruyter.

**Stehouwer, Herman / Durco, Matej / Auer, Eric Auer / Broeder, Daan** (2012): "Federated Search: Towards a Common Search Infrastructure", in: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*: 3255-3259.

**Wierlacher, Alois** (1997): „Verfehlte Alterität. Zum Diskurs deutschsprachiger Reiseführer über fremde Speisen“, in: Teuteberg, Hans Jürgen / Neumann, Gerhard / Wierlacher, Alois (eds.): *Essen und kulturelle Identität*. Europäische Perspektiven. Berlin: Akademie Verlag 498-509.