Carolina Flinz

Tourlex: ein deutsch-italienisches Fachwörterbuch zur Tourismussprache für italienische DaF-Lerner

1	Einleitung	2.6	Phase der Vorbereitung für das Onlinerelease
2	Die Phasen des lexikographischen Prozesses bei Tourlex	3	Zusammenfassung und Ausblick
2.1	Vorbereitungsphase	4	Literatur
2.2	Phase der Datenbeschaffung	4.1	Aufsätze und Monographien
2.3	Phase der Computerisierung	4.2	Zitierte Wörterbücher
2.4	Phase der Datenaufbereitung	5	Bildverzeichnis
2.5	Phase der Datenauswertung		

Abstract: Tourlex is a specialized bilingual online dictionary under construction hosted at the University of Mannheim with a particular focus on collocations and multi-word units. The languages included are German and Italian, but because of the need for online dictionaries of tourism language (Flinz 2015: 56) the framework is open to the inclusion of other languages. Tourlex is a corpus-based dictionary, i.e. the primary sources will be corpora, in particular a proper bilingual comparable corpus analysed with the tools Sketch Engine and Lexpa, and the freely accessible corpus DEREKO. The aim of this paper is to give an overview of the main actions (already done but also in planning), according to the phases of the lexicographical process of a dictionary under construction. The description of each phase will be enriched by examples taken from the project, showing also how the decisions taken to satisfy the needs of the user, the Italian learner of German as a foreign language, had influenced the microstructure of the entries. We conclude with a final reflection on the data, facts, and ongoing problems.

Keywords: bilingual dictionary, specialized dictionary, language of tourism, dictionary of German as a foreign language, Italian-German

Schlagwörter: Bilinguales Wörterbuch, Fachwörterbuch, Tourismussprache, DaF-Wörterbuch, Italienisch-Deutsch

Carolina Flinz, Universität Pisa, Dipartimento di Economia e Management, Lungarno Pacinotti, 43, 56126, Pisa, PI, Italien, e-Mail: c.flinz@ec.unipi.it

<https://doi.org/10.1515/lexi-2018-0003>

1 Einleitung

Tourlex¹ ist ein zweisprachiges Internet-Fachwörterbuch zur Tourismussprache (Deutsch-Italienisch) für italienische DaF-Lerner, die Tourismuswissenschaft studieren. Es handelt sich dabei um ein neu konzipiertes Ausbauwörterbuch² (Schröder 1997: 60), dessen Primärquellen Korpora sind, die auf der Basis von computergestützten Analysetools (u. a. Sketch Engine und Lexpa) untersucht werden. Das Wörterbuch, das sich sowohl an Lernende (Studenten der Tourismuswissenschaft) als auch an sonstige im Tourismusbereich Tätige richtet, wird als Pilotprojekt vorerst Stichwörter aus einem besonders komplizierten Bereich („die Allgemeinen Reisebedingungen“) enthalten und einen besonderen Fokus auf Kollokationen und usuelle Wortverbindungen (Steyer 2013: 14) legen, die Fremdsprachenlernern große Probleme bereiten (Hausmann 1984: 395). Diese Schwierigkeiten können nämlich weder mit anderen Tourismus-Fachwörterbüchern (vgl. Flinz 2015: 56) noch mit den üblichen mehrsprachigen Ressourcen gemeistert werden: Grund dafür ist, dass diese Ressourcen relevante Begriffe entweder nicht als Lemma aufführen oder, wenn sie vorhanden sind, kaum Informationen zu den fachtypischen Wortverbindungen bieten.

Ziel dieses Beitrags ist es, einen Überblick über die Phasen des lexikographischen Prozesses bei Tourlex zu geben, ein Prozess, der aber noch nicht abgeschlossen ist, wie es bei dynamischen Wörterbüchern der Fall ist (Klosa 2016: 29), und von einem ständigen Überdenken charakterisiert ist. Da Internetwörterbücher untereinander sehr unterschiedlich³ sind und diese Unterschiede auch die zugrundeliegenden lexikographischen Prozesse beeinflussen,⁴ wurde das Modell von Klosa⁵ verwendet, die explizit auf Wörterbücher im Ausbau Bezug nimmt. Hauptmerkmale des lexikographischen Prozesses bei Tourlex, wie bei anderen Internetwörterbüchern, sind das Zusammenfließen und sich Überlappen der Phasen wie in einem Kreis (Klosa 2013: 518), denn die Veröffentlichung der Daten beginnt, noch bevor das Wörterbuch beendet ist. Die Bearbeitung der Stichwörter folgt anderen Krite-

1 Das Projekt Tourlex wurde dank eines Forschungsstipendiums für erfahrene Wissenschaftler der Alexander von Humboldt Stiftung an der Universität Mannheim (Lehrstuhl von Frau Prof. Storrer) ermöglicht.

2 Wörterbücher, die ständig erweitert, ausgebaut und veröffentlicht werden, auch wenn sie nicht ‚vollständig‘ sind, werden auch „dynamische Wörterbücher“ (Lemberg 2001: 81) oder „Wörterbücher im Aufbau“ (Storrer/Freese 1996: 117; Storrer 2001: 65) genannt.

3 Für eine Klassifizierung von Internetwörterbüchern vgl. u. a. Engelberg/Storrer (2016).

4 Erste Auseinandersetzungen und Überlegungen zu computerlexikographischen Prozessen für die Erstellung von Internetwörterbüchern sind erst ab 2003 zu finden; vgl. u. a. Müller-Spitzer (2003); Derouin/Le Meur (2008); Klosa/Tiberius (2016); Hildenbrandt/Klosa (2016).

5 Klosa (2013: 517f) erweitert Wiegands 5-Phasenmodell (Wiegand 1998: 135), das für gedruckte Wörterbücher konzipiert wurde, und problematisiert als erste den unterschiedlichen Verlauf der Phasen, die nicht mehr hintereinander, sondern parallel ablaufen.

rien als der alphabetischen Reihenfolge, wie zum Beispiel Modulen (Haß 2005: 13 f.), der Frequenz oder Wortartenklassen (Storrer 2001: 63).⁶ In Tourlex wird vorerst die Wortart „Substantiv“ bearbeitet und die Module werden nach Frequenzschichten behandelt.

Im Folgenden werden die wichtigsten Arbeitsschritte des lexikographischen Prozesses vorgestellt und thematisiert, wie man der Grafik in Abbildung 1 entnehmen kann.

Mit diesem Beitrag soll auch veranschaulicht werden, wie wichtig der Kontakt mit den Benutzern eines Wörterbuches (in diesem Fall mit italophonen DaF-Lernern) in allen Phasen des Prozesses ist. Feedback sowie Äußerung von Problemfällen seitens der Rezipienten sind im Arbeitsverlauf berücksichtigt worden: Da das Wörterbuch für unterschiedliche Ziele, wie u. a. für die Rezeption, Produktion und die Übersetzung von Texten, benutzt werden soll, waren die Problemfälle nicht nur für die Auswahl der Angaben in der Mikrostruktur des Wörterbuchartikels von großer Bedeutung, sondern auch für ihre Organisation. Zwei unterschiedliche Varianten der Mikrostruktur wurden seitens der DaF-Studierenden der Tourismusfakultät und der DaF-Lernenden einer Tourismusfachschule bewertet, so dass die endgültige Auswahl als Ausdruck der tatsächlichen Bedürfnisse der DaF-Lerner entstanden ist.

Tourismus ist ein wichtiger Wirtschaftsbereich der Europäischen Union und in vielen Ländern bieten Universitäten Studiengänge in Tourismuswissenschaften an. Mehrsprachige wissenschaftliche Fachwörterbücher zur Tourismussprache könnten wegen ihres hohen Bedarfs großes Potential auch im Fremdsprachenunterricht haben, da Wörterbücher Einfluss auf den Fremdsprachenerwerb haben (Domínguez Vázquez/Mollica/Nied Curcio 2014: 2; Giacomini 2014: 222).⁷ Tourlex könnte im DaF-Unterricht benutzt werden, nicht nur um den Wortschatz des Tourismus leichter zugänglich zu machen, sondern auch um den Erwerb der richtigen Kollokationen und usuellen Wortverbindungen zu fördern, die die Produktion von Fachtexten erschweren.

Problemfälle und Baustellen des Projektes werden in diesem Beitrag ebenfalls Berücksichtigung finden und Teile eines Wörterbuchartikels werden als Exemplifizierung des Prozesses dargestellt. Daten, Fakten und Ausblicke werden abschließend thematisiert und diskutiert.

⁶ Vorteile der unterschiedlichen Kriterien werden u. a. in Storrer (2001: 63) erläutert.

⁷ Vgl. u. a. die Studien von Nesi 2000 und Laufer 2000.

Abb. 1: Mind-Map des lexikographischen Prozesses bei Tourlex.⁸

⁸ Die Mind-Map wurde mit dem Programm MindMup2 erstellt (vgl. <https://www.mindmup.com>, letzter Zugriff am 21. 1. 2018).

2 Die Phasen des lexikographischen Prozesses bei Tourlex

2.1 Vorbereitungsphase

In der Vorbereitungsphase wurde viel Zeit in die inhaltliche Konzeption des Wörterbuches⁹ und die Exploration der computertechnischen Möglichkeiten investiert. Erste Ideen zur inhaltlichen Konzeption sowie darauffolgende Entscheidungen bezüglich der äußeren und inneren Zugriffsstrukturen¹⁰ waren erst nach unterschiedlichen Analysen möglich:

- a) Eine Bestandsaufnahme der existierenden Internetwörterbücher zur Tourismussprache für einen ersten Überblick über die involvierten Sprachen, die Suchmöglichkeiten, die Verweisstrukturen und über multimediale Eigenschaften der lexikographischen Ressourcen (vgl. Flinz 2015: 43–56). Durch die Analyse konnte festgestellt werden, dass die existierenden Fachwörterbücher zur Tourismussprache selten das Sprachenpaar Deutsch-Italienisch beinhalten und dass nur ein Drittel der analysierten Wörterbücher über Suchfunktionen sowie Verweise verfügt. Die Mikrostruktur enthielt meistens nur die Äquivalenzangabe, während Grammatik- und Beispielangaben selten vorhanden und Ausspracheangaben als Audiodatei in keinem einzigen Wörterbuch enthalten sind;
- b) Umfragen, um die Bedürfnisse und hypothetischen Benutzungssituationen der anvisierten Benutzer feststellen zu können, und Umfragen im Sinnes eines *simultaneous feedback* (in Anlehnung an de Schryver/Prinsloo 2000: 199). Beide waren für das Prozedere der Arbeit von großer Bedeutung.

2014 wurden 80 Studenten der Tourismuswissenschaft (Universität Pisa) mit Hilfe eines Online-Fragebogens zu folgenden Kernfrageblöcken interviewt: Wörterbuchbenutzer, Wörterbuchtyp, Wörterbuchbenutzungssituation, Bedürfnisse und Wörterbuchformat. Die Ergebnisse (Flinz 2014a: 13–33) haben sowohl die Einstellungen der Probanden gegenüber unterschiedlichen Wörterbuchtypen und Formaten beleuchtet, als auch ihre persönliche Bevorzugung für bestimmte makro-, medio- und mikrostrukturelle Eigenschaften hervorgehoben; somit konnten ein erster Eindruck von den möglichen Bedürfnissen der hypothetischen Benutzer gewonnen und erste Hypothesen über die möglichen Ziele des Fachwörterbuches aufgestellt werden.

Zwei weitere Umfragen wurden 2017 durchgeführt: hier waren die Probanden erneut 50 Studenten der Tourismuswissenschaft, aber auch 39 Schüler (13. Klasse) einer Tourismusfachschule in Pisa (Istituto Antonio Pacinotti). Die Fragen haben

⁹ Erste Überlegungen zum Projekt haben schon 2014 (Flinz 2014: 323–334) stattgefunden, aber die ursprüngliche Konzeption hat sich aufgrund der Korpuserstellung und der empirischen Arbeit mit den Korpora deutlich verändert und eine komplett neue Richtung bekommen.

¹⁰ Zu diesen typischen Begriffen der Internetlexikografie vgl. u. a. Engelberg/Müller-Spitzer/Schmidt (2016: 154–160)

sich nicht nur an den oben genannten Frageblöcken orientiert, sondern haben auch zwei alternative Muster von Wörterbuchartikeln zur Auswahl gestellt. Konkret ging es um die Organisation der Informationen im Angabenteil der „Typischen Wortverbindungen“ vom Lemma *Buchung*: In Variante A waren die Belege direkt nach der Übersetzung der Kollokation, in Variante B am Ende des Wörterbuchartikels:

Variante A:¹¹

Buchung, die

- Aggettivo – Sostantivo :
eine kurzfristige **Buchung: una prenotazione Last-Minute; una prenotazione a breve termine**
eine telefonische **Buchung: una prenotazione telefonica**
- Sostantivo – Verbo:
eine **Buchung** vornehmen: **effettuare una prenotazione**

Esempi:

Der Reisegast haftet gegenüber Camino-Reisen für alle Verpflichtungen von Mitreisenden, für die er die **Buchung** vornimmt, aus dem Reisevertrag. (Urscher 2017)

L'intermediario presso il quale *sia stata effettuata la prenotazione* del pacchetto turistico non risponde in alcun caso delle obbligazioni nascenti dall'organizzazione del viaggio. (Eden 2017)

eine **Buchung** bestätigen: **confermare una prenotazione**

Esempi:

Sonderwünsche sowie **Buchungen** unter einer Bedingung und mündliche Nebenabreden sind nur dann gültig, wenn sie von INTER CHALET schriftlich bestätigt werden. (Interchalet 2017)

Lo stesso importo verrà applicato in caso di qualsiasi ulteriore richiesta da parte del turista per modifiche a *prenotazione* già avvenuta e *confermata*. (Viaggi del Turchese 2017)

Variante B:

Buchung, die

- Aggettivo – Sostantivo :
eine kurzfristige **Buchung: una prenotazione Last-Minute; una prenotazione a breve termine**
eine telefonische **Buchung: una prenotazione telefonica**

¹¹ Hier wird nur ein kleiner Ausschnitt aus dem Wörterbuchartikel als Exemplifizierung der zwei Varianten vorgestellt. Die Probanden haben hingegen einen Screenshot von zwei tatsächlichen Probestartikeln bekommen.

– Sostantivo – Verbo :

eine **Buchung** vornehmen: **effettuare una prenotazione**

eine **Buchung** bestätigen: **confermare una prenotazione**

Esempi:

Der Reisegast haftet gegenüber Camino-Reisen für alle Verpflichtungen von Mitreisenden, für die er die **Buchung** vornimmt, aus dem Reisevertrag. (Urscher 2017)

L'intermediario presso il quale *sia stata effettuata la prenotazione* del pacchetto turistico non risponde in alcun caso delle obbligazioni nascenti dall'organizzazione del viaggio. (Eden 2017)

Sonderwünsche sowie **Buchungen** unter einer Bedingung und mündliche Nebenabreden sind nur dann gültig, wenn sie von INTER CHALET schriftlich bestätigt werden. (Interchalet 2017)

Lo stesso importo verrà applicato in caso di qualsiasi ulteriore richiesta da parte del turista per modifiche a *prenotazione* già avvenuta e *confermata*. (Viaggi del Turchese 2017)

Die Ergebnisse haben gezeigt, dass ca. 80 % der Probanden die Variante A bevorzugt haben. Einige Begründungen waren u. a. folgende:

1. „Perché così strutturato il dizionario risulta più facilmente leggibile oltre al fatto che ogni possibile combinazione del vocabolo ricercato risulta meglio inserita in una frase o in un contesto generico con l'esempio indicato subito sotto“ (Prob. 1);
2. „Intuitivo, diretto“ (Prob. 2);
3. „Perché sono allegati esempi che ti fanno comprendere“ (Prob. 3);
4. „Ho riscontrato che con gli esempi si capisce al meglio l'utilizzo della parola“ (Prob. 4);
5. „Lo trovo migliore perchè mi sembra più pratico, facile e veloce da consultare“ (Prob. 6);
6. „Perché risulta, a mio parere, più comprensibile. Il mio schema mentale individua più facilmente vocaboli soprattutto negli esempi. E' uno schema ordinato e l'esempio subito sotto alla spiegazione permette di fissare al meglio il concetto“ (Prob. 8).¹²

12 Die deutsche Übersetzung lautet: 1. „So strukturiert ist das Wörterbuch besser lesbar und jede mögliche Kombination mit dem gesuchten Lemma kann innerhalb eines Satzes oder eines Kontextes dank des darauffolgenden Beispiels eingesehen werden“; 2. „intuitiv, direkt“; 3. „Beispiele unterstützen das Verstehen“; 4. „Ich habe bemerkt, dass man mit Beispielen den Gebrauch des Wortes besser verstehen kann“; 5. „Ich finde es besser, weil man praktischer, einfacher und schneller die gewünschten Informationen suchen kann“; 6. „Meiner Ansicht nach ist es besser verständlich. Die Lexeme können in den Beispielen schneller identifiziert werden. Und das Beispiel direkt unter der Bedeutung unterstützt auch das Lernen“.

Die Ergebnisse haben gezeigt, dass alle analysierten lexikographischen Ressourcen Kollokationen, usuelle Wortverbindungen und Phraseologismen aufweisen, meistens aber ohne Systematik. Beispielsätze werden nicht immer zu den richtigen Bedeutungen hinzugefügt und, wie man schon vorausahnen kann, waren kaum Kollokationen aus dem Fachsprachebereich Tourismus vorhanden. Für Studierende der Tourismuswissenschaft, die Deutsch als Fremdsprache lernen, sind syntagmatische Angaben und insbesondere Kollokationen mit Exemplifizierungen unumgänglich. Ihre systematische Organisation im Angabenteil erleichtert die Auffindung der Informationen und unterstützt den Wortschatz- und Kollokationserwerb.

Die geschilderten Analysen, die auch als Marktanalysen und Nutzerstudien definiert werden können, haben dem Projekt einen großen Impuls gegeben und haben sich auch auf die inhaltliche Planung ausgewirkt. Ein Organisationsplan mit Angaben zum Zeitplan und Workflow musste entsprechend angepasst werden.

In dieser Phase haben auch Überlegungen zur digitalen Umgebung stattgefunden: Die ursprüngliche Idee, die Freeware Glossword¹⁵ zu benutzen, wurde aufgrund der ungeklärten Nachhaltigkeit ausgeschlossen. Nach Abwägung unterschiedlicher Möglichkeiten wurde das Mediawiki-System ausgewählt, das als System von Wiktionary und Wikipedia weit verbreitet und bekannt ist.¹⁶ Unterschiedliche Vorteile sind mit diesem System verbunden:

- Man kann direkt online arbeiten; Vorlagen für oft benutzte Textabschnitte sind vorhanden.
- Die Strukturen können hierarchisch aufgebaut werden.
- Es ist leicht expandierbar (sowohl was die Zahl der Einträge als die involvierten Sprachen betrifft).
- Eine Suchfunktion mit eingabe- und volltextbasierter Suchmöglichkeit ist integriert.
- Audiodateien in unterschiedlichen Formaten (u. a. mp3, mp4) sind leicht einsetzbar.
- Die Versionierung der Artikel sowie das Anzeigen der letzten Änderungen werden vom System unterstützt.

Das System ermöglicht auch die Verwaltung der Benutzerrechte, die zur Zeit nur an einzelne Benutzer vergeben worden sind; eine Differenzierung der Benutzerrechte im Sinne eines halb-kollaborativen Wörterbuches ist angedacht.

Durch das Hosting und die technische Unterstützung der Universität Mannheim wurde somit auch die Nachhaltigkeit des Projekts gewährleistet.¹⁷

¹⁵ Vgl. <http://glossword.biz> (letzter Zugriff am 1.1.18).

¹⁶ Dies ist auch aus der Benutzerbefragung hervorgegangen: Wiktionary wird von den italienischen DaF-Lernern benutzt.

¹⁷ Die Vorteile eines XML-Editors mit einem damit verbundenen Redaktionssystem waren mir natürlich bewusst, aber nach Bewertung der daraus entstehenden Kosten (für die Programmierung und die nachträgliche Betreuung) fiel eine Entscheidung dagegen aus.

2.2 Phase der Datenbeschaffung

In der Phase der Datenbeschaffung wurden die Quellen¹⁸ für die Wörterbuchbasis des Projektes zusammengestellt und die Tondateien für die Ausspracheangaben im Deutschen aufgenommen. Da es sich um ein Ausbauwörterbuch handelt, wird sich diese Phase bis zum Ende des Projektes hinziehen. Als Primärquellen wurden Textkorpora herangezogen, während als sekundäre Quellen andere Wörterbücher verwendet wurden, sowohl im Print- als auch im Online-Format; als tertiäre Quellen wurden andere sprachliche Quellen, wie u. a. Grammatiken und Handbücher, benutzt.

Die Tondateien mussten alle gesondert aufgenommen werden und wurden, da die native Aussprache eines Lemmas garantiert werden muss, von einer deutschen Muttersprachlerin gesprochen.¹⁹ Audiodateien sind für Fremdsprachenlerner besonders wichtig, da das mehrmalige Abhören der vertonten Aussprache auch den Lerneffekt des Fachsprachenvokabulars erhöht.

2.2.1 Primärquellen

Lemnitzer und Zinsmeister (2015: 137) raten, sich folgende Frage zu stellen, wenn man mit Korpora arbeitet: „Selber kochen oder auswärts essen gehen?“. Die intuitive, aber auch von den Autoren begründete Antwort ist die zweite: „auswärts essen gehen“, d. h. bestehende Korpora (wie z. B. DEREKo²⁰ und das DWDS-Korpus²¹) zu verwenden, die u. a. die Kriterien der Größe und Repräsentativität realisieren und auch frei zugänglich sind. Leider war diese Möglichkeit für Tourlex für die fachsprachliche Seite nicht möglich, denn touristische Texte konnten nicht ohne Weiteres von den großen, frei verfügbaren Korpora isoliert werden, weder um eine provisorische Lemmakandidatenliste zu erstellen, noch um die fachsprachlichen Kollokationen herauszufiltern. Die Entscheidung fiel auf ein kombiniertes Vorgehen:

Erstens wurde ein vergleichbares deutsch-italienisches Spezialkorpus²² auf der Basis einer Teiltextsorte des Tourismusbereichs in Sketch Engine erstellt (die „Allgemeinen Reisebedingungen“²³ und „Condizioni generali di contratto di vendita di pacchetto turistico“):

¹⁸ Zu Quellen (Primär-, Sekundär-, Tertiärquellen) in der Lexikographie vgl. Wiegand (1998: 140).

¹⁹ Ein besonderer Dank geht an zwei Mitarbeiterinnen vom Lehrstuhl von Frau Prof. Storrer, die mich sowohl bei der Aufnahme als auch beim Hochladen der Tondateien unterstützt haben.

²⁰ Vgl. <http://www1.ids-mannheim.de/kl/projekte/korpora/>.

²¹ Vgl. <https://www.dwds.de/r>.

²² Ein Korpus, das eine Fachsprache zum Gegenstand hat, kann als Spezialkorpus definiert werden (Lemnitzer/Zinsmeister 2015: 141).

²³ Folgende Reiseveranstalter wurden berücksichtigt: Camino-Reisen, Dertour, FTI Touristik, Interchalet, Kiwi Tours, Novasol, Olimar, Piccolonia, Reisegeier, TUI Wolters für das deutsche Korpus und Alpitour, Columbia Turismo, Eden Viaggi, I Viaggi del Turchese, Italia Tourism Online, Itermar, King Holidays, Settemari, Veratour, Viaggi dell'Elefante.

Abb. 3: Ausschnitt der häufigsten Wortformen (absolute Frequenz) des Vergleichskorpus.²⁴

Das Vergleichskorpus hat jeweils 40.000 Token pro Sprache, aber als dynamisches Korpus (oder auch Monitorkorpus) kann es erweitert und ausgebaut werden. Es wurden Metadaten (u. a. Autor des Textes, Textsorte, Datum, Link zur Online-Quelle) erstellt und, da das Korpus zur Verfügung gestellt werden soll, wurde eine Klärung der Urheberrechte eingeleitet.

Das ad hoc erstellte Vergleichskorpus wurde benutzt, um die provisorische Lemmakandidatenliste (absolute und relative Häufigkeit auf der Basis der integrierten Referenzkorpora German Web 2013 und Italian Web 2016) zu erstellen, um die Äquivalente zu determinieren und die Kollokationen sowie usuellen Wortverbindungen zu extrahieren (vgl. dazu Abschnitt 2.4.2).

In einem zweiten Schritt wurden auch die oben genannten Korpora (DEReKo und das DWDS-Korpus) in die Arbeit integriert, um weitere Beispiele und Kontexte der Stichwörter herauszufiltern. DEReKo wurde insbesondere als Referenzkorpus für die deutschsprachigen Angaben sowohl für Erstellung der Stichwortliste als auch für das Herausfiltern der syntagmatischen Muster für die Bestimmung von Kollokationen und usuellen Wortverbindungen (vgl. DEReKOLL) verwendet.²⁵

Es wurde ein korpusbasierter quantitativ-qualitativer Ansatz (Lemnitzer/Zinsmeister 2015: 37) gewählt, da die Daten aus Korpora abgeleitet, aber zusätzlich interpretiert und lexikographisch bearbeitet werden.

²⁴ Im deutschen Korpus: *Reise* (276), *Kunde* (165), *Reisepreis* (161), *Olimar* (141), *Buchung* (139), *Anspruch* (136), *Reisevertrag* (132), *Reiseantritt* (129), *Leistung* (128), *Reisebeginn* (119), *Änderung* (96). Im italienischen Korpus: *organizzatore* (236, deutsche Übersetzung ‚Veranstalter‘), *turista* (226, deutsche Übersetzung ‚Tourist‘), *contratto* (197, deutsche Übersetzung ‚Vertrag‘), *pacchetto* (197, deutsche Übersetzung ‚Pauschalreise‘), *viaggio* (160, deutsche Übersetzung ‚Reise‘), *art.* (157, deutsche Übersetzung ‚Artikel‘), *partenza* (136, deutsche Übersetzung ‚Reisebeginn‘), *servizi* (125, deutsche Übersetzung ‚Leistungen‘), *prezzo* (119, deutsche Übersetzung ‚Preis‘), *catalogo* (78, deutsche Übersetzung ‚Katalog‘), *prenotazione* (63, deutsche Übersetzung ‚Buchung‘).

²⁵ Ich möchte an dieser Stelle ein herzliches Dankeschön an den Programmbereich Korpuslinguistik des Instituts für Deutsche Sprache aussprechen und insbesondere an Rainer Perkuhn, der mich im Laufe des Projektes vielfach unterstützt hat.

2.2.2 Sekundär- und Tertiärquellen

Als Sekundärquellen wurden sowohl Wörterbücher im Printformat als auch im Online-Format verwendet:

- das DWDS-Wörterbuch, canoo.net, Duden Online und Wiktionary für grammatische Informationen (Genus, Numerus, Deklination), IPA-Zeichen und Worttrennung;
- Il nuovo dizionario di tedesco, Leo, dict.cc, Langenscheidt, Pons, Langenscheidt Tourismus, Wiktionary, um Äquivalente, syntagmatische Angaben und Beispiele zu untersuchen.

Die Einträge wurden dort aufgesucht und wenn sie enthalten waren, wurden die Angaben verglichen und gegebenenfalls entnommen. Als Tertiärquellen wurden Handbücher und Grammatiken zu Rate gezogen. Zu dieser dritten Kategorie werden auch die sprachliche Intuition des Lexikographen und die linguistische Bibliothek (Geyken/Lemnitzer 2016: 200 f.) hinzugerechnet, die auch in diesem Projekt eine wichtige Rolle spielen.

2.3 Phase der Computerisierung

Für das Projekt hat keine spezifische Programmierung von Korpusrecherchertools stattgefunden, sondern es wurden bekannte und getestete Tools ausgewählt, wie Sketch Engine, Lexpan und COSMAS II:

- Mit Hilfe von Sketch Engine²⁶ (vgl. Kilgariff et al. 2004: 108–112) wurden Wortlisten, Kookkurrenzen und Wortprofile auf der Basis von syntaktischen Kriterien ermittelt;
- Mit Lexpan (Lexical Pattern Analyzer, vgl. Steyer 2013: 110 f.)²⁷ wurden syntagmatische Strukturen auf Festigkeit, Varianz, Slotbesetzung und kontextuelle Einbettungsmuster explorativ untersucht;
- Mit COSMAS II (vgl. Bodmer Mory 2014) wurde das Deutsche Referenzkorpus²⁸ (DeReKo) abgefragt.

26 Sketch Engine ist ein kommerzielles Tool, das u. a. auf morphosyntaktischen Annotationen in der jeweiligen Sprache operiert. Für das Deutsche wurde für die Erarbeitung des Wortprofils (Word Sketch) das empfohlene German RFTagger 4.1 benutzt und für das Italienische das Marco Baroni Tagset 2.0.

27 Lexpan ist ein einzelsprachenunabhängiges Analyseprogramm, das im Projekt „Usuelle Wortverbindungen“ des Instituts für Deutsche Sprache entwickelt wurde (<http://www1.ids-mannheim.de/lexik/uvw/lexpan.html>). Ein besonderes Dankeschön geht an Kathrin Steyer für die wichtigen Ratschläge zur Auffindung von Kollokationen und usuellen Wortverbindungen.

28 DeReKo ist das Korpusrecherche- und Analysesystem (Version 2.2.1) des Instituts für Deutsche Sprache. Es enthält 42 Milliarden Wörter (Stand 03.02.2018): <http://www1.ids-mannheim.de/kl/projekte/korpora.html>

Die digitale Aufbereitung der Texte (Tagging und Lemmatisierung) ist entfallen, da Sketch Engine die Korpora automatisch syntaktisch parst sowie annotiert. Für Lexpan ist diese Prozedur nicht nötig, da im Programm Suchanfragen über KWIC-Zeilen definiert werden. Trotzdem hat aber eine Aufbereitung und Bereinigung der Texte stattgefunden, denn die Texte mussten von PDF- ins Text-Format konvertiert werden und im Korpusabfragesystem hochgeladen werden.

Für Tourlex wurde auf die Einrichtung eines lexikographischen Redaktions-systems im üblichen Sinne verzichtet. Es wurde aber trotzdem eine redaktionelle Arbeitsumgebung definiert, da die Funktionalitäten des Wiki-Systems zur Datensicherung und Versionierung benutzt worden sind. Die Wörterbuchinhalte werden in einem relationalen Datenbankverwaltungssystem gespeichert, das MySQL benutzt. Das System wird vom Rechenzentrum der Universität Mannheim programmiert, das auch in bestimmten Zeitabschnitten Backups zur Sicherung der Daten vornimmt.

Es wurde auch ein einfaches elektronisches Redaktionshandbuch erstellt, das inhaltliche redaktionelle Richtlinien mit Beispielen zur Artikelstruktur beinhaltet. Auf eine konsistente Datenstruktur bei allen Wörterbuchartikeln wird viel Wert gelegt, so dass auch am Ende des Projekts die Informationen weitergeleitet werden können.

2.4 Phase der Datenaufbereitung

In der Phase der Datenaufbereitung haben zwei große Arbeitsschritte stattgefunden: 1. die Erstellung der provisorischen Lemmakandidatenliste und 2. die Modellierung der lexikographischen Daten in einer Datenbankstruktur.

2.4.1 Die provisorische Lemmakandidatenliste

Zur Erstellung der Stichwortkandidatenliste wurde die automatische Korpusextraktion des Korpusabfragesystems benutzt; anschließend wurde auch ein formaler Korrekturabgleich mit anderen Wörterbüchern (sofern es möglich war) durchgeführt. Es wurde folgendermaßen vorgegangen:

Erstens wurden auf der Basis der Funktion „Word List“, dem automatischen Lemmatisierer von Sketch Engine, Flexionsformen Wortformen zugewiesen, so dass die Grundformen des Wortes gedeutet werden konnten.²⁹ Die daraus resultierende Liste (1013 Stichwörter), die auf der Basis der absoluten Frequenz erstellt wurde, wurde anschließend extrahiert und auf Substantive reduziert (vgl. Tabelle 1).

Diese Liste wurde auch anschließend benutzt, um die Frequenzschichten zu erstellen, die als Grundlage für die zu bearbeitenden Module dienten.

²⁹ Die Option „lemma“ und Mindestfrequenz „5“ wurden ausgewählt.

Tab. 1: Mit Sketch Engine erstellte Lemmaliste (nur Substantive).

	Stichwort	Frequenz im Korpus
1	<i>Reise</i>	266
2	<i>Buchung</i>	139
3	<i>Leistung</i>	128
4	<i>Veranstalter</i>	113
5	<i>Änderung</i>	96
6	<i>Rücktritt</i>	83
7	<i>Ziffer</i>	76
8	<i>Schaden</i>	69
9	<i>Zahlung</i>	69
10	<i>BGB</i>	66

Tab. 2: Beispiel aus der Extraktions-Liste mit Sketch Engine.

Single-word	Score	F	Reff
<i>Olimar</i>	2.850,60	141	840
<i>Reisevertrag</i>	2.025,49	131	7.191
<i>Reisebeginn</i>	1.514,64	119	12.991
<i>Reiseantritt</i>	1.352,31	129	20.014
<i>Reisegast</i>	1.338,83	80	5.142
<i>Reisepreis</i>	1.199,07	160	35.892
<i>Reiseleistung</i>	1.156,57	74	6.909
<i>Inter-Chalet</i>	904,79	48	2.352
<i>Piccolonia</i>	885,83	42	0
<i>FTI</i>	794,40	53	8.058

In einem zweiten automatischen Arbeitsschritt wurden mit der Funktion ‚Keywords und Terms‘ die Keywords des deutschen Korpus im Vergleich zum integrierten Referenzkorpus (German Web 2013) erarbeitet (vgl. Tabelle 2).

Nach Herausfilterung der Eigennamen der Veranstalter, wie u. a. *Olimar*, *Inter-Chalet*, *Piccolonia*, konnten die signifikantesten Schlüsselwörter des Korpus determiniert werden: *Reisevertrag*, *Reisebeginn*, *Reiseantritt*, *Reisegast*, *Reisepreis*, *Reiseleistung* etc.

Durch den Vergleich mit dem Referenzkorpus DeReKo (2017-II) wurde zusätzlich eine Rangliste der Schlüsselwörter des deutschen Korpus (allerdings auf der Basis von Wortformen) aufgrund von unterschiedlichen Assoziationsmaßen (Chi-Quadrat und Log-Likelihood-Ratio) erstellt (vgl. Tabelle 3).

30 Informationen zu den Abkürzungen: Rg.: Rang; DeR: DeReKo; RB: Korpus Reisebedingungen; Win: Winner.

Tab. 3: Die Term-Extraktions-Liste mit DeReKo als Referenzkorpus.³¹

Rg.	DeR	RB	win	chi2	llr	
1.	5	42	RB	27.167.967,25	1.101,55	Piccolonia
2.	79	48	RB	13.132.074,68	1.126,91	CHALET
3.	1.423	141	RB	9.201.230,95	2.857,91	Olimar
4.	610	80	RB	6.713.953,03	1.663,85	Reisegast
5.	2	9	RB	5.330.262,93	232,43	Reiseendes
6.	0	7	RB	5.067.043,18	188,89	XFTI
7.	4	9	RB	4.510.219,71	226,82	ticketlosem
8.	0	6	RB	4.343.179,87	161,91	Gruppenauftraggeber
9.	2	6	RB	3.257.381,90	152,91	XTUI
10.	0	4	RB	2.895.453,24	107,94	Bausteinreiseleistungen

Tab. 4: Informationen zu Okkurrenzen und Streuung der Lemmata in DeReKo und in den Reisebedingungen.

	DeReKo	Reisebedingungen	Streuung
Gruppenauftraggeber	0	6	1
Bausteinreiseleistungen	0	4	1

Auch aus dieser Liste wurden die Eigennamen (wie u. a. *Piccolonia*, *Chalet*, *Olimar*, *XFTI*, *XTUI*) oder andere, nicht auf den Tourismus bezogene Wörter, wie *Hyperlink* etc. entfernt. Im Vergleich zum Referenzkorpus in Sketch Engine sind jedoch andere Wörter in den höheren Reihen des Rankings zu finden: u. a. *Reisegast*, *Reiseende*; *Gruppenauftraggeber*, *Bausteinreiseleistung*, *Pauschalreiseleistung*.³¹

Anschließend wurde auch die Streuung der Wörter in den unterschiedlichen Subkorpora (die einzelnen Reiseveranstalter) verifiziert. Interessant waren neben den durchgängig belegten Stichwörtern u. a. die in Tabelle 4 erfassten Fälle.

Die Wörter *Gruppenauftraggeber* und *Bausteinreiseleistungen* waren kein einziges Mal in DeReKo zu finden, sondern hatten jeweils 6 bzw. 4 Okkurrenzen im Vergleichskorpus, jedoch nur in einem einzigen Katalog (Urscher bzw. FTI). Die Erklärung dafür ist folgende:

- Urscher ist ein Reiseveranstalter, der Busreisen organisiert und mit Gruppen arbeitet. Die Figur des Gruppenauftraggebers ist für diese Art von Reisen sehr wichtig, denn er ist für die ganze Gruppe verantwortlich. Somit hat sich das Wort als wichtiger Kandidat für die Stichwortliste ergeben;
- FTI ist ein Reiseveranstalter, der dem Kunden die Möglichkeit bietet, bestimmte Reiseleistungen nach Reisetypen oder besonderen Wünschen zusammenzu-

³¹ Exemplarisch wurde hier die Anordnung nach dem Assoziationsmaß chi2 dargestellt.

stellen. Die Reise kann aus Leistungen bestehen, die wie in einem Baukasten zusammengestellt werden. Da diese Art von Reisen zwar neu, aber doch immer mehr verbreitet ist, wurde entschieden, auch dieses Wort als Lemma hinzuzufügen.

Die provisorische Stichwortliste³² wurde also aus dem Zusammenspiel dieser drei Arbeitsschritte erstellt: Berechnung der absoluten Häufigkeit der Lemmakandidaten im Vergleichskorpus; Erstellung einer Rangliste der Lemmakandidaten auf der Basis der relativen Häufigkeit in Sketch Engine; Erstellung einer Rangliste der Lemmakandidaten auf der Basis von DEREKo mit zusätzlicher Berücksichtigung der Streuung in den jeweiligen Katalogen. Ein korrektiver formaler Vergleich mit anderen Wörterbuchlisten konnte nicht durchgeführt werden, da keine anderen Wörterbücher dieses Typus zum Sprachenpaar Deutsch-Italienisch existieren.

2.4.2 Die Modellierung der lexikographischen Daten

Auf der Basis der provisorischen Lemmakandidatenliste hat die Erarbeitung der lexikographischen Daten begonnen, die dann in einer entsprechenden Datenbank modelliert wurden. Als Lösung wurde Excel verwendet; die Daten wurden dann exportiert und im relationalen Datenbankverwaltungssystem von Media-Wiki gespeichert. Für die Erstellung der lexikographischen Daten wurden folgende Arbeitsschritte vorgenommen:

Erstens wurden Informationen zur Grammatik (Genus, Numerus, Deklination), Orthographie (Silbentrennung) und Aussprache (Transkription in IPA-Zeichen) in anderen Wörterbüchern untersucht. Wie man aus dem Diagramm in Abb. 4 entnehmen kann, konnten nur einige Informationen aus anderen Quellen entnommen werden:

1. 100 % der grammatischen Angaben (Genus, Numerus, Deklination) aus dem Wörterbuch Canoo.net;
2. 71 % der orthographischen Angaben (Worttrennung) aus dem Wörterbuchportal DWDS;
3. 50 % der Ausspracheangaben (IPA-Transkription) aus Wiktionary.³³

Die größten Probleme haben sich bei der Transkription der Lemmata in IPA-Zeichen ergeben. Hierzu musste die Hälfte der Angaben neu erstellt werden unter Zuhilfenahme von Printwörterbüchern, wie dem Aussprache-Duden, oder Tertiärquellen, wie Handbüchern.

Im zweiten Schritt wurden die Äquivalente in existierenden mehrsprachigen Internetwörterbüchern (Leo, dict.cc, Pons Online, Langenscheidt, Wiktionary, Lan-

³² Die provisorische Stichwortliste besteht aus 175 Lemmata, die bis zum Ende des Projektes komplett online gestellt werden.

³³ Alle entnommenen Informationen wurden nachgeprüft und gegebenenfalls korrigiert.

Abb. 4: Quellen für die Angaben zur Grammatik und Zeichengestalt.

Abb. 5: Vorhandensein von Äquivalenten in Sekundärquellen.

genscheidt Tourismus) verifiziert. Die Situation kann als wie in Abb. 5 gezeigt dargestellt werden.

Etwa die Hälfte der Lemmata der provisorischen Stichwortliste ist nicht als Eintrag in anderen Wörterbüchern verzeichnet. Auch wenn sie dort vorkommen, gibt

es meistens keinen Vermerk zur fachsprachlichen Variante aus dem Tourismusbereich, wie u. a. folgende Beispiele hervorheben:

- Die Übersetzung von *Buchung* mit *registrazione* in dict.cc: *registrazione* ist nur mit der Bedeutung von Buchung als ‚das Verbuchen von Belegen auf Konten in der Buchführung‘³⁴ zusammenzubringen und nicht als ‚Reservierung‘, dies würde *prenotazione* entsprechen;
- Die Übersetzung von *Umbuchung* mit *scrittura di storno* in Leo: *scrittura di storno* hat nichts mit dem Tourismus zu tun, sondern wird mit der Bedeutung aus der Wirtschaftssprache ‚an eine andere Stelle im Konto oder auf ein anderes Konto buchen‘ übersetzt (die Variante aus dem Tourismusbereich wäre *modifica di prenotazione* oder *variazione di prenotazione*);
- Die Übersetzung von *Zielgebiet* in Leo und Zanichelli mit *obiettivo di attacco* oder *zona di attacco* ist eine Variante aus dem Militärbereich ‚Gebiet, Fläche als Ziel militärischer Angriffe‘, anstelle der aus dem Tourismus *meta turistica* oder *località di soggiorno* etc.³⁵

Die Äquivalente wurden also mithilfe des Vergleichskorpus identifiziert: Ohne Rückschlüsse aus den daraus entnommenen syntaktischen, semantischen, morphologischen und lexikalischen Beziehungen wäre ihre Identifizierung nicht möglich gewesen.³⁶ Alle 175 Stichwörter der provisorischen Stichwortkandidatenliste wurden schließlich mit einem oder mehreren Äquivalenten versehen und die Äquivalente in italienischer Sprache mit Genusangaben ergänzt.

Im letzten Arbeitsschritt wurden Kookkurrenzen mit Beurteilung ihres Status als Kandidaten für Kollokationen extrahiert (signifikantes Vorkommen der Einheiten, aber in syntaktischer Beziehung) und usuelle Wortverbindungen (usualisierte Einheiten des Lexikons) wurden induktiv rekonstruiert und auf Festigkeit, Varianz und Musterhaftigkeit (Steyer 2013: 76) überprüft. Hierzu wurden unterschiedliche Schritte unternommen:

- a) Mit der Funktion Word-Sketch, die automatisch ein „Wort-Profil“ auf der Basis von morphosyntaktischen Kriterien erstellt, konnten für das Korpus relevante Kollokationen sowohl in italienischer als auch in deutscher Sprache ermittelt werden.
- b) Mit Lexpan konnte eine induktive Rekonstruktion von Wortverbindungen und Wortverbindungsmodellen von sprachlichen Oberflächen ausgehend (Steyer

³⁴ Vgl. <https://www.duden.de/rechtschreibung/Buchung#Bedeutung1>

³⁵ Diese Tatsache hat erneut den Bedarf von zweisprachigen touristischen Fachwörterbüchern im DaF-Unterricht hervorgehoben: ohne sie können einfache Textverstehens- oder Übersetzungsaktivitäten nicht gelöst werden, da die spezifischen Äquivalente nicht in den anderen Wörterbüchern enthalten sind.

³⁶ Zu der Notwendigkeit von Vergleichskorpora für die Identifizierung von Äquivalenten vgl. u. a. Sinclair/Payne/Pérez 1996: 177; Calzolari 1996: 6; Prinsloo 2013: 1346.

2013: 17) in beiden Sprachen erfolgen. Die KWIC-Konkordanzen wurden aus Sketch Engine importiert und für Lexpan in sogenannte „Stores“ konvertiert, die anschließend hochgeladen und durchsucht wurden. Die Erstellung von Füllertabellen, die die Frequenzen der lexikalischen Lückenfüller auflisten sowie ihre unterschiedlichen Sortierungsmöglichkeiten anzeigen, war von großem Nutzen für die Rekonstruktion von Wortverbindungen.

- c) Durch die Nutzung von DEREKOLL für Tourlex konnten die syntagmatischen Muster der Lemmata der provisorischen Stichwortliste aus CCDB und DEREKo extrahiert werden. Mit diesem Instrument konnten die spezifischen Kollokationen des Spezialkorpus denen der Standardsprache gegenübergestellt werden³⁷.

Um die Daten vergleichen zu können, mussten alle Felder auch in unterschiedlichen Kombinationen analysiert und für jedes Lemma die Informationen kontrastiv zusammengefasst werden (vgl. Tabelle 5).

Es muss hinzugefügt werden, dass die gesamte Phase von einer manuellen Analyse der Korpusdaten bestimmt wurde, auch wenn sie computergestützt war. Die Informationen aus anderen Wörterbüchern wurden „konservativ“ und nicht automatisch extrahiert.

³⁷ Es sind interessante Fälle zum Vorschein gekommen, wie u. a. die bevorzugten Verbindungen des Lexem *Rücktritts*, die durch Rücktritte bestimmter Personen von Positionen in der Politik oder Sport geprägt sind und nicht auf Reisekontexte hinweisen (vgl. Flinz/Perkuhn 2018).

Tab. 5: Kollokationen und usuelle Wortverbindungen für das Lemma *Buchung*.

	Adjektiv-Substantiv	Verb-Substantiv	Präposition-Substantiv	Andere
Deutsch	eine kurzfristige Buchung	eine Buchung vornehmen	bei Buchung	von der Buchung zurücktreten
Italienisch	<i>una prenotazione a breve termine</i>	<i>effettuare una prenotazione</i>	<i>all'atto della prenotazione; al momento della prenotazione</i>	<i>recedere dalla prenotazione/ dal contratto</i>
Übersetzung	eine kurzfristige Buchung	eine Buchung vornehmen	bei Buchung	von der Buchung/vom Vertrag zurücktreten
Deutsch	eine telefonische Buchung	eine Buchung bestätigen	vor Buchung	
Italienisch	<i>una prenotazione telefonica</i>	<i>confermare una prenotazione</i>	<i>prima della prenotazione; prima della conclusione del contratto</i>	
Übersetzung	eine telefonische Buchung	eine Buchung bestätigen	vor Buchung; vor dem Abschluss des Vertrags	
Deutsch			nach Buchung	
Italienisch			<i>successivamente alla prenotazione/ alla stipula del contratto; dopo la conclusione della prenotazione</i>	
Übersetzung			nach Buchung; nach dem Abschluss des Vertrages; nach Abschluss der Buchung	

2.5 Phase der Datenauswertung

In der Phase der Datenauswertung sind sowohl (teil-)automatisch-korpuslinguistische sowie redaktionell-lexikographische Arbeiten durchgeführt worden. Nicht nur die automatisch erstellte Stichwortliste wurde redaktionell überprüft (vgl. Abschnitt 2.2.1), sondern auch die Belege, die automatisch aus dem Vergleichskorpus ermittelt wurden. Der Eingang der Belege in die Wörterbuchartikel hat stets nach einer formalen Überarbeitung stattgefunden, um die Qualität der Einträge gewährleisten zu können (Klosa 2010: 157). Auch die Auswahl ist nicht automatisch erfolgt, sondern manuell unter Berücksichtigung von Kriterien, wie u. a. Variation und Länge (z. B. nicht mehr als ein Absatz).

Eine der Hauptaufgaben dieser Phase war zudem das Verfassen neuer Wörterbuchartikel, die kontinuierlich bis zum Ende des lexikographischen Prozesses stattfinden wird. Für jeden Artikel wurden Korpusdaten in Form von KWICs, Wortprofilen, Tabellen mit Slots und Füllern aus dem Vergleichskorpus analysiert und interpretiert, Informationen aus den anderen Sekundärquellen wurden entnommen, Korpusbelege wurden ausgewählt und angepasst.

In dieser Phase wurde auch die Vernetzungsstruktur³⁸ des Wörterbuches erstellt, d. h. die inhaltlichen Beziehungen zwischen den Wörtern wurden durch Verweise zwischen den einzelnen Wörterbuchartikeln (vgl. Engelberg/Lemnitzer 2009: 179) hergestellt; Tabelle 6 gibt einen Überblick über diese Beziehungen ausgehend vom Lemma *Buchung*.

Die Stichwörter wurden auf der Basis von unterschiedlichen Kriterien verlinkt – synonymische Beziehungen (Angabe „Sinonimi“) und semantische Nähe aufgrund des Kookkurrenzprofils (Angabe „Parole collegiate“).

Als Grundlage wurde das Korpus benutzt und zur Verifizierung der Synonymrelationen wurde sowohl die Funktion „Sketch Diff“ von Sketch Engine als auch die Option „contrast near-synonyms“ der CCDB³⁹ verwendet, die eine Kontrastierung (des allgemeinsprachlichen Gebrauchs) von Beinahe-Synonymen ermöglicht.

Für die Auswahl der naheliegenden Wörter wurde erneut vom eigenerstellten Korpus ausgegangen (Funktion „Thesaurus“ von Sketch Engine); zusätzlich wurden

Tab. 6: Erarbeitung von Vernetzungen des Lemmas *Buchung*.

Lemma	Thesaurus	CCDB	Tourlex
Buchung	Reisevertrag, Reise, Reiseleistung, Kunde, Rücktritt, Reisebedingung, Umbuchung	Reisebüro, Reservierung, Reisen, Bestellung, Umbuchung, Reiseveranstalter	Kunde, Hotel, Reise, Reisebedingung, Reisebüro, Rücktritt, Umbuchung

³⁸ An dieser Stelle möchte ich mich herzlich bei Herrn Prof. Schierholz bedanken, der mich auf diese Problematik bei einem Vortrag in Erlangen hingewiesen hat.

³⁹ Zur Kookkurrenzdatenbank (CCDB) vgl. u. a. Belica (1995) und Belica (2007).

die Vernetzungen auch in der Kookkurrenzdatenbank⁴⁰ („semantic proximity model“) untersucht.

Die Einträge wurden weiterhin mit externen Quellen vernetzt: für jeden Eintrag wurden Links zum entsprechenden Eintrag im DWDS-Wortauskunftssystem,⁴¹ in Canoo.net⁴² und OWID (das Online-Wortschatz-Informationssystem Deutsch),⁴³ wenn die Lemmata dort enthalten waren, hinzugefügt. Diese Entscheidung wurde aufgrund der Bedürfnisse der Adressaten des Wörterbuches getroffen, die aus den Fragebögen hervorgegangen sind.

In dieser Phase wurden auch die Tondateien, die in der Datenbeschaffungsphase aufgenommen worden waren, erneut angehört und getestet, um eventuelle Problemfälle zu identifizieren und ggf. neue Aufnahmen zu machen.

2.6 Phase der Vorbereitung für das Onlinerelease

In der Vorbereitung für das Onlinerelease sind sowohl technische wie inhaltliche Arbeitsschritte vorgesehen. Die technischen Schritte haben insbesondere mit der Freischaltung der Wörterbuchartikel im Internet zu tun und sind zu Teilen vom Rechenzentrum der Universität Mannheim durchgeführt worden. In dieser Phase wurden auch Entscheidungen zu den Zugriffsstrukturen getroffen (d. h. wie der Zugriff auf das Wörterbuch von außen ermöglicht wird⁴⁴): Außer der eingabe- und filterbasierten Suche, die im System Mediawiki integriert ist, wurde entschieden, eine zusätzliche indexbasierte Suche zu ermöglichen: nach der alphabetischen Leiste wurden die Lemmata mit den italienischen Äquivalenten (vgl. Abb. 6) verlinkt:

Stichwörter

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

A [Bearbeiten]

- Abänderung (variazione; modifica)
- Abhilfe (rimedio)
- Abhilfeverlangen (richiesta di rimedio)
- Abs. (paragrafo, comma)
- Abschluss (conclusione)
- Änderung (modifica; variazione)

Abb. 6: Screenshot der alphabetischen Leiste in Tourlex.

⁴⁰ Vgl. <http://corpora.ids-mannheim.de/ccdb/> (letzter Zugriff 01.03.18).

⁴¹ <https://www.dwds.de>

⁴² <http://www.canoo.net>

⁴³ <http://www.owid.de>

⁴⁴ Für einen Überblick über die Zugriffsstrukturen in Internetwörterbüchern vgl. Engelberg/Müller-Spitzer/Schmidt (2016: 153 f.).

Buchung

Buchung, die

Inhaltsverzeichnis [Verbergen]	
1	Grammatica
2	Ortografia
3	Fonetica
4	Traduzione/
5	Combinazioni tipiche
6	Sinonimi
7	Derivazione e/o composizione
8	Parole collegate
9	Link

Grammatica [\[Bearbeiten\]](#)

Sostantivo (femminile)

	Singolare	Plurale
Nominativo	die Buchung	die Buchungen
Genitivo	der Buchung	der Buchungen
Dativo	der Buchung	den Buchungen
Accusativo	die Buchung	die Buchungen

Ortografia [\[Bearbeiten\]](#)

Suddivisione in sillabe: Bu-chung

Abb. 7: Screenshot des Lemmas *Buchung* in Tourlex (Angaben zu Grammatik und Orthographie).

Der Entwurf der Artikel wurde außerdem getestet, um zu sehen, wie die Angaben online erscheinen würden, und ggf. Änderungen vorzunehmen. Nach dem Feedback von ca. 50 Benutzern wurde die Variante bevorzugt, die die unterschiedlichen Bedürfnisse der DaF-Lerner an der Tourismus fakultät und -fachschule am meisten befriedigen kann:

- die grammatikalischen Angaben wurden in Tabellenform aufgelistet (vgl. Abb. 7), um visuell das Flexionsparadigma des Substantivs darzustellen, und somit ihren Erwerb zu unterstützen.
- Die Aussprache wurde in ihrer vertonten Variante vorangestellt, so dass die Benutzer zuerst das Wort hören und dann lesen können (vgl. Abb. 8): das Aktivieren mehrerer Sinneskanäle begünstigt die Speicherung der Informationen im Langzeitgedächtnis und fördert somit auch den Fremdspracherwerb. Da viele dieser Fachwörter Komposita sind, ist die Aussprache für Fremdsprachener lerner sehr schwierig; die Audiodatei und die Transkription ermöglichen ein schnelles Erlernen der normkonformen Aussprache;

Fonetica [Bearbeiten]Pronuncia: die Buchung

Trascrizione fonetica: /bu:χʊŋ/

Abb. 8: Screenshot des Lemmas *Buchung* in Tourlex (Angaben zur Zeichengestalt).**telefonische Buchung:** prenotazione telefonica

Esempi:

Telefonische Buchungen, welche kürzer als 7 Tage vor Reisebeginn erfolgen, sind für den Kunden verbindlich. (Piccolonia 2018)

Con la prenotazione telefonica, elettronica o cartacea, a CASAMUNDO viene assegnato in maniera vincolante un incarico di mediazione turistica. (Casamundo 2018)

Abb. 9: Screenshot des Lemmas *Buchung* in Tourlex (Kollokationen und Beispiele).

- c) Kollokationen und usuelle Wortverbindungen wurden in mehreren Kategorien systematisiert (Adjektiv/Substantiv; Substantiv/Verb; Präposition/Verb und weitere), so dass auch induktiv über die möglichen Verbindungen des Fachwortes im Unterricht reflektiert werden kann. Mit der graphischen Hervorhebung des Lemmas (fett und schwarz) und der Übersetzung der Kollokation (fett und grau) wird auch die visuelle Wahrnehmung der syntagmatischen Verbindung unterstützt.

Beispiele aus dem deutschen und italienischen Korpus (vgl. Abb. 9), die durch die Hervorhebung des Wortes „Esempi“ (rot und fett) eingeleitet werden, folgen der Wortverbindung direkt, so dass die Exemplifizierung der Kollokationen den Lernprozess erleichtern kann.

Die Entscheidung, Belege sowohl aus dem deutschen als auch dem italienischen Korpus hinzuzufügen, wurde gemäß den Benutzerrückmeldungen getroffen: Mit diesem „doppelten“ Beleg können die Benutzer von authentischen Quellen für unterschiedliche Zielsetzungen profitieren. Die direkte Verknüpfung mit dem Korpus bedarf noch infrastruktureller und nutzungsrechtlicher Klärung.

In dieser Phase wurde ebenfalls die Benutzeroberfläche getestet, Hyperlinks und Suchmöglichkeiten wurden erprobt. Die Wörterbuchumtexte (wie die Benutzungshinweise, Informationen zum Inhalt und Korpus etc.) wurden erstellt und hochgeladen.

3 Zusammenfassung und Ausblick

Tourlex ist ein Pilotprojekt, und da ein Ausbau des Wörterbuches angestrebt wird, ist die Offenlegung des lexikographischen Prozesses von großer Bedeutung, sodass daraus auch zukünftige Kooperationen oder Projekte profitieren können.

Eine Erweiterung des Wörterbuches ist in mehrere Richtungen angedacht:

- Ausweitung des Korpus mit zusätzlichen Texten der Teiltextsorte „Reisebedingungen“ sowohl in deutscher als auch in italienischer Sprache;
- Ausweitung des Korpus durch Hinzufügung von anderen Textsorten aus dem Tourismusbereich in deutscher und italienischer Sprache (u. a. Reiseprospekte und Reiseführer);
- Ausweitung des Korpus auf die entgegengesetzte Richtung und auf andere Sprachen. Auch wenn der primäre Benutzer der italienische Muttersprachler ist, der im Tourismus arbeitet, Deutsch als Fremdsprache lernt oder gelernt hat und sich mit deutschsprachigen Touristen beschäftigt, sind trotzdem Entscheidungen für die Benutzung des Wörterbuches aus der Sicht von deutschen Muttersprachlern (u. a. die Hinzufügung des Genus für das italienische Äquivalent und Beispiele in beiden Sprachen) getroffen worden.

Im Verlauf des Projekts sind weitere Schritte geplant, die eine noch nähere Anpassung an tatsächliche Benutzerbedürfnisse anstreben, wie u. a. ein Experiment zum Testen des Wörterbuches mit italienischen Probanden und die Erarbeitung eines Online-Fragebogens, der von den Benutzern heruntergeladen werden kann. Eine stärkere Mitwirkung im Sinne eines halbkollaborativen Wörterbuches ist hingegen noch in Abwägung.

Tourlex kann den Benutzern schon in seiner jetzigen Form in unterschiedlicher Weise behilflich sein: Es kann sie bei der Rezeption von Texten und beim Verständnis von Fachwörtern, aber auch bei der Produktion von Fachtexten und der Übersetzung von Fachwörtern (täglicher Schriftverkehr etc.) unterstützen. Und da Wörterbücher „des Wortschatzlernens einzige Rettung“ sind (Hausmann 1993: 480), kann Tourlex auch eine konkrete Unterstützung für den DaF-Unterricht an Tourismusfakultäten und Tourismusfachschulen sein.

4 Literatur

4.1 Aufsätze und Monographien

- Belica, Cyril (1995): *Statistische Kollokationsanalyse und Clustering. COSMAS-Korpusanalysemodul*. Mannheim: Institut für Deutsche Sprache. [Unter: <http://corpora.ids-mannheim.de/ccdb/>; letzter Zugriff: 08. 03. 18].
- Belica, Cyril (2007): *Koikkurrenzdatenbank CCDB – V3. Eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs*. Mannheim: Institut für Deutsche Sprache.
- Bodmer Mory, Franck (2014): Mit COSMAS II „in den Weiten der IDS-Korpora unterwegs“. In: *Ansichten und Einsichten. 50 Jahre Institut für Deutsche Sprache*. Mannheim: Institut für Deutsche Sprache, 376–385.
- Calzolari, Nicoletta (1996): Lexicon and corpus: a Multi-faceted Interaction. In: Gellerstam, Martin et al. (Hrsg.): *EURALEX '96 Proceedings I*. Göteborg: Göteborg University, 3–16.

- Derouin, Marie-Jeanne/Le Meur, Andre (2008): ISO-Standards for Lexicography and Dictionary Publishing. In: Bernal, Elisenda/De Cesaris, Janet (Hrsg.), *Proceedings of the 13th EURALEX International Congress*. Barcelona: Documenta Universitaria, 663–668.
- Domínguez Vázquez, María José (2015): Wörterbuchbenutzung: Tendenzen, riskante Entwicklungen und aktuelle Fragestellungen an die Lexikographie. In: Calañas Continente, José-Antonio/Robles i Sabater, Ferran (Hrsg.), *Die Wörterbücher des Deutschen: Entwicklungen und Perspektiven*. Frankfurt a. M.: Lang, 269–302.
- Domínguez Vázquez, María José/Mollica, Fabio/Nied Curcio, Martina (2014): Zweisprachige Wörterbücher, Didaktik und Translation: Einführung. In: Domínguez Vázquez, María José/Mollica, Fabio/Nied Curcio, Martina (Hrsg.), *Zweisprachige Lexikographie zwischen Translation und Didaktik*. Berlin/Boston: de Gruyter, 1–12.
- Engelberg, Stefan/Lemnitzer, Lothar (2009): *Lexikographie und Wörterbuchbenutzung*. 4. überarb. u. erw. Aufl. Tübingen: Stauffenburg.
- Engelberg, Stefan/Müller-Spitzer, Carolin/Schmidt, Thomas (2016): Vernetzungs- und Zugriffsstrukturen. In: Klosa, Annette/Müller-Spitzer, Carolin (Hrsg.): *Internetlexikografie. Ein Kompendium*. Berlin/Boston: de Gruyter, 153–195.
- Engelberg, Stefan/Storrer, Angelika (2016): Typologie von Internetwörterbüchern und -portalen. In: Klosa, Annette/Müller-Spitzer, Carolin (Hrsg.): *Internetlexikografie. Ein Kompendium*. Berlin/Boston: de Gruyter, 31–63.
- Flinz, Carolina (2014): TOURLEX: erste Bausteine für ein deutschitalienisches Lexikon der Touristik-Fachsprache. In: Domínguez Vázquez, María José/Mollica, Fabio/Nied Curcio, Martina (Hrsg.), *Zweisprachige Lexikographie zwischen Translation und Didaktik*. (= Lexicographica. Series Maior 147) Berlin/Boston: de Gruyter, 323–338.
- Flinz, Carolina (2014a): Wörterbuchbenutzung: Ergebnisse einer Umfrage unter Studenten der Tourismuswissenschaft. In: Mann, Michael (Hrsg.): *Digitale Lexikographie. Ein- und mehrsprachige elektronische Wörterbücher mit Deutsch: aktuelle Entwicklungen und Analysen*. Hildesheim u. a.: Olms, 13–33.
- Flinz, Carolina (2015): Online-Fachwörterbücher des Tourismus: Bestandsaufnahme und Darstellung der Hauptmerkmale. In: Calañas Continente, José A./Robles i Sabater, Ferran (Hrsg.): *Die Wörterbücher des Deutschen. Entwicklungen und Perspektiven*. Frankfurt a. M.: Peter Lang, 43–58.
- Flinz, Carolina/Perkuhn, Rainer (2018): Wortschatz und Kollokationen in „Allgemeine Reisebedingungen“. Eine intralinguale und interlinguale Studie zum fachsprachlich-lexikographischen Projekt „Tourlex“. In: Krek, Simon et al. (Hrsg.): *Proceedings of the 18th Euralex International Congress, 17–21 July 2018, Ljubljana, Slovenia*, 959–966.
- Geyken, Alexander/Lemnitzer, Lothar (2016): Automatische Gewinnung von lexikografischen Angaben. In: Klosa, Annette/Müller-Spitzer, Carolin (Hrsg.), *Internetlexikografie: Ein Kompendium*. Berlin/Boston: de Gruyter, 195–241.
- Giacoma, Luisa (2014): Wie würde ein Wörterbuch aussehen, wenn der Benutzer es selbst schreiben könnte? In: Domínguez Vázquez, María José/Mollica, Fabio/Nied Curcio, Martina (Hrsg.), *Zweisprachige Lexikographie zwischen Translation und Didaktik*. Berlin/Boston: de Gruyter, 221–246.
- Haß, Ulrike (Hrsg.) (2005): *Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz*. (= Schriften des Instituts für Deutsche Sprache 12). Berlin/New York: de Gruyter.
- Hausmann, Franz J. (1984): Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortverbindungen. In: *Praxis des neusprachlichen Unterrichts* 31, 385–406.
- Hausmann, Franz J. (1993): Ist der deutsche Wortschatz lernbar? In: *Informationen Deutsch als Fremdsprache* 20, 471–485.
- Hildenbrandt, Vera/Klosa, Annette (Hrsg.) (2016): Lexikographische Prozesse bei Internetwörterbüchern. In: OPAL 01/2016.

- Institut für Deutsche Sprache (2017): *Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2017-I* (Release vom 08. 03. 2017). Mannheim: Institut für Deutsche Sprache. PID: 10932/00-0373-23CD-C58F-FF01-3.
- Kilgariff, A. et al. (2004): The Sketch Engine. In: Williams, G./Vessier, S. (Hrsg.): *Proceedings of the 11th EURALEX International Congress, Lorient, France, July 6–10*. Bd. 1. Lorient: Université de Bretagne Sud, 105–115.
- Klosa, Annette (2010): On the combination of automated information and lexicographically interpreted information in two German online dictionaries. In: Granger, Sylviane/Paquot, Magali (Hrsg.): *eLexicography in the 21st century. New challenges, new applications. Proceedings of eLex 2009. Louvain-la-Neuve, 22–24 October 2009*. (Cahiers du Cental 7) Louvain: UCL Presses Universitaires, 157–163.
- Klosa, Annette (2013): The lexicographical process (with special focus on online dictionaries). In: Gouws, Rufus et al. (Hrsg.): *Dictionaries. An International Encyclopedia of Lexicography*. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlin u. a.: de Gruyter Mouton, 517–524.
- Klosa, Annette (2016): Der lexicographische Prozess im Projekt elexiko. In: Hildenbrandt, Vera/Klosa, Annette (Hrsg.): *Lexikographische Prozesse bei Internetwörterbüchern*. OPAL 1/2016 Mannheim: Institut für Deutsche Sprache, 29–38.
- Klosa, Annette/Tiberius, Carole (2016): Der lexicographische Prozess. In: Klosa, Annette/Müller-Spitzer, Carolin (Hrsg.): *Internetlexikografie. Ein Kompendium*. Berlin/Boston: de Gruyter, 65–110.
- Lauffer, Batia (2000): Electronic dictionaries and incidental vocabulary acquisition: does technology make a difference? In: Heid, Ulrich et al. (Hrsg.): *Proceedings of the Ninth EURALEX International Congress*. Stuttgart: Institut für Maschinelle Sprachverarbeitung, 849–854.
- Lemberg, Ingrid (2001): Aspekte der Online-Lexikographie für wissenschaftliche Wörterbücher. In: Lemberg, Ingrid/Schröder, Bernd/Storrer, Angelika (Hrsg.): *Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher*. Tübingen: Niemeyer, 71–91.
- Lemnitzer, Lothar/Zinsmeister, Heike (2015): *Korpuslinguistik. Eine Einführung*. 3. Aufl. Tübingen: Narr.
- Müller-Spitzer, Carolin (2003): Ordnende Betrachtungen zu elektronischen Wörterbüchern und lexicographischen Prozessen. In: *Lexicographica* 19, 140–168.
- Nesi, Hilary (2000): Electronic Dictionaries in Second Language Vocabulary Comprehension and Acquisition: the State of the Art. In: Heid, Ulrich et al. (Hrsg.): *Proceedings of the Ninth EURALEX International Congress*. Stuttgart: Institut für Maschinelle Sprachverarbeitung, 839–847.
- Prinsloo, D. J. (2013): The utilization of bilingual corpora for the creation of bilingual dictionaries. In: Gouws, Rufus H. et al. (Hrsg.): *Dictionaries. An International Encyclopedia of Lexicography*. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography Berlin/Boston: de Gruyter Mouton, 1344–1356.
- Schröder, Martin (1997): Brauchen wir ein neues Wörterbuchkartell? Zu den Perspektiven einer computerunterstützten Dialektlexikographie und eines Projekts „Deutsches Dialektwörterbuch“. In: *Zeitschrift für Dialektologie und Linguistik* 64, 1, 57–66.
- de Schryver, Gilles-Maurice/Prinsloo, D. J. (2000): Dictionary Making Process with ‘Simultaneous Feedback’ from the Target Users to the Compilers. In: Heid, Ulrich et al. (Hrsg.): *Proceedings of the 9th EURALEX International Congress*, Stuttgart: Institut für Maschinelle Sprachverarbeitung, 197–209.
- Sinclair, John/Payne, Jonathan/Pérez Hernández, Chantal (Hrsg.) (1996): Corpus to Corpus: a study of Translation Equivalence. In: *International Journal of Lexicography* 9 (3), 171–178.
- Steyer, Kathrin (2013): *Usuelle Wortverbindungen. Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht*. Tübingen: Narr.

- Storrer, Angelika (2001): Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertext-konzepts in der Lexikographie. In: Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika (Hrsg.): *Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher*. Tübingen: Niemeyer, 53–69.
- Storrer, Angelika/Freese, Katrin (1996): Wörterbücher im Internet. In: *Deutsche Sprache* 24, 97–136.
- Wiegand, Herbert Ernst (1998): *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. 1. Teilband. Berlin/ New York: de Gruyter.
- Wodak, Ruth (2015): ‚Normalisierung nach rechts‘: Politischer Diskurs im Spannungsfeld von Neoliberalismus, Populismus und kritischer Öffentlichkeit. In: *Linguistik Online* 7 (4)4, 27–44. [Unter: <https://bop.unibe.ch/linguistik-online/article/view/2191/3355>; letzter Zugriff: 08.03.18]

4.2 Zitierte Wörterbücher

*Canoo*net = *Deutsche Wörterbücher und Grammatik* <<http://www.canoo.net>>
Dict.cc = *Dizionario tedesco-italiano* <<https://deit.dict.cc>>
DUDEN = <<https://www.duden.de>>
DWDS = *Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart*. <<https://www.dwds.de>>
Zanichelli = *Il nuovo dizionario di tedesco*. Hrsg. v. Giacomina, Luisa/Kolb, Susanne Bologna: Zanichelli/Klett Pons, 3. Aufl. 2014.
Langenscheidt = *Deutsch-Italienisch Wörterbuch* <<https://de.langenscheidt.com/deutsch-italienisch>>
Langenscheidt Fachwörterbuch = *Tourismus* <<https://woerterbuch.langenscheidt.de/home/produkte/italienisch-deutsch-tourismus-fachwoerterbuch>>
Leo = *Wörterbücher Deutsch, Italienisch* <<https://www.leo.org/italienisch-deutsch/>>
OWID = *Online-Wortschatz-Informationssystem Deutsch* <<http://www.owid.de>>
PONS = *Online-Wörterbuch* <<https://de.pons.com>>
Wiktionary = <<https://www.wiktionary.org>>

5 Bildverzeichnis

- Abb. 1: Mind-Map des lexikographischen Prozesses bei Tourlex
- Abb. 2: Untersuchte Stichwörter aus der provisorischen Stichwortkandidatenliste
- Abb. 3: Ausschnitt der häufigsten Wortformen (absolute Frequenz) des Vergleichskorpus
- Abb. 4: Quellen für die Angaben zur Grammatik und Zeichengestalt
- Abb. 5: Vorhandensein von Äquivalenten in Sekundärquellen
- Abb. 6: Screenshot der alphabetischen Leiste in Tourlex
- Abb. 7: Screenshot des Lemmas *Buchung* in Tourlex (Angaben zu Grammatik und Orthographie)
- Abb. 8: Screenshot des Lemmas *Buchung* in Tourlex (Angaben zur Zeichengestalt)
- Abb. 9: Screenshot des Lemmas *Buchung* in Tourlex (Kollokationen und Beispiele)