

Herausforderungen bei der Erstellung der multilingualen, korpusbasierten lexikografischen Ressource CombiDigiLex¹

Manuel FERNÁNDEZ MÉNDEZ

Universidade de Santiago de Compostela (España)

manuel.fernandez.mendez@usc.es

<https://orcid.org/0000-0002-8217-5298>

Inmaculada MAS ÁLVAREZ (Autorin für die Korrespondenz)

Universidade de Santiago de Compostela (España)

inmaculada.mas@usc.es

<https://orcid.org/0000-0002-5387-7282>

Meike MELISS

Universidade de Santiago de Compostela (España)

meike.meliss@usc.es

<https://orcid.org/0000-0003-0380-8905>

Abstract: Der Beitrag stellt die methodologischen und technologischen Herausforderungen bei der Erstellung der multilingualen, korpusbasierten lexikografischen Ressource CombiDigiLex (<https://combidigilex.wixsite.com/deutsch>) dar und verfolgt dabei hauptsächlich drei Ziele. Zunächst werden der anwendungsorientierter Forschungskontext und die Gründe für die Entwicklung des mehrsprachigen, lexikalisch-grammatischen Informationssystems CombiDigiLex, welches für die Sprachen Deutsch, Spanisch und Portugiesisch Information für den Einsatz im L2-Produktions- und Übersetzungsprozess bereitstellen soll, dargestellt. Weiterhin werden die relevantesten Merkmale der besagten Ressource in ihrer aktuellen Version (v.1.0.6) vorgestellt. Abschließend werden die methodologischen und technischen Schwierigkeiten der Datengewinnung sowie der lexikografischen Datenmodellierung und Darstellung, die nicht nur während der offiziellen Projektlaufzeit überwunden werden mussten, sondern die selbst zum Zeitpunkt der Abfassung dieses Artikels noch teilweise offene Fronten bilden skizziert.

Schlüsselwörter: Lexikografie; kontrastive Linguistik; Fremdsprachenerwerb; Korpuslinguistik

English:

Challenges in the creation of the multilingual corpus-based lexicographic resource CombiDigiLex

¹ Dieser Artikel steht in direktem Zusammenhang mit dem Forschungsprojekt COMBIDIGILEX, das von den Professorinnen Mas Álvarez und Meliss zwischen 2016 und 2019 an der Universidade de Santiago de Compostela (Spanien) koordiniert wurde. Das Projekt wurde vom Ministerium für Wirtschaft und Wettbewerbsfähigkeit und vom Europäischen Fonds für regionale Entwicklung finanziert (Referenz: FFI2015-64476-P). Siehe auch unsere Projektseite in drei Sprachen: <https://combidigilex.wixsite.com/deutsch> [letzter Abruf: 27.09.2021].



Abstract: The paper presents the methodological and technological challenges in the creation of the multilingual, corpus-based lexicographic resource CombiDigiLex (<https://combidigilex.wixsite.com/deutsch>), with three main aims. First, the application-oriented research context and grounds for the development of the multilingual lexico-grammatical information system CombiDigiLex, which is intended to provide information for the languages German, Spanish and Portuguese, for its use in the L2 production and translation process, are presented. Furthermore, the most relevant features of the mentioned resource in its current version (v.1.0.6) are presented. Finally, the methodological and technical difficulties of data gathering as well as lexicographic data displaying, that not only had to be overcome during the official project period, but which even at the time of writing this paper are still partly open fronts, are outlined.

Keywords: lexicography; comparative linguistics; second language learning; corpus linguistics

Español:

Retos en la elaboración del recurso lexicográfico multilingüe basado en corpus lingüísticos CombiDigiLex

Resumen: El artículo presenta los retos metodológicos y tecnológicos que los autores han experimentado en la creación del recurso lexicográfico multilingüe basado en corpus CombiDigiLex (<https://combidigilex.wixsite.com/website-1>), con tres objetivos principales. En primer lugar, se presenta el contexto de investigación, orientada a la aplicación, y la justificación del desarrollo del sistema de información léxico-gramatical multilingüe CombiDigiLex, cuyo objetivo es proporcionar información para las lenguas alemana, española y portuguesa, para su uso en el proceso de producción y traducción de L2. Además, se presentan las características más relevantes de dicho recurso en su versión actual (v.1.0.6). Por último, se exponen las dificultades metodológicas y técnicas de la obtención de datos, así como de la configuración y presentación de la información lexicográfica, que no solo hubo que superar durante el periodo oficial del proyecto, sino que incluso en el momento de redactar este documento siguen siendo en parte frentes abiertos.

Palabras clave: lexicografía; lingüística comparada; aprendizaje de segundas lenguas; lingüística de corpus

Català:

Reptes en l'elaboració del recurs lexicogràfic multilingüe basat en corpus lingüístics CombiDigiLex

Resum: L'article presenta els reptes metodològics i tecnològics que els autors han experimentat en la creació del recurs lexicogràfic multilingüe basat en corpus CombiDigiLex (<https://combidigilex.wixsite.com/deutsch>), amb tres objectius principals. En primer lloc, es presenta el context de la investigació, orientada a l'aplicació, i la justificació del desenvolupament del sistema d'informació lexicogramatical multilingüe CombiDigiLex, l'objectiu del qual és proporcionar informació per a les llengües alemanya, castellana i portuguesa, per al seu ús en el procés de producció i traducció de segones llengües. Alhora, es presenten les característiques més rellevants del mencionat recurs en la seva versió actual (v.1.0.6). Finalment, s'exposen les dificultats metodològiques i tècniques de l'obtenció de dades, i també de la configuració i presentació de la informació lexicogràfica, que es van haver de superar durant el període oficial del projecte i que, fins i tot en el moment de redactar aquest article, continuen sent en part fronts oberts.

Paraules clau: lexicografia; lingüística comparada; aprenentatge de segones llengües; lingüística de corpus

Einleitung

Der Beitrag verfolgt hauptsächlich folgende drei Ziele: (i) Darstellung des Forschungskontextes und der Gründe für die Entwicklung des mehrsprachigen, lexikalisch-grammatischen Informationssystems CombiDigiLex (<https://combidigilex.wixsite.com/deutsch>), welches für die Sprachen Deutsch, Spanisch und Portugiesisch korpusbasierte Information für den Einsatz im L2-Produktions- und Übersetzungsprozess bereitstellen soll; (ii) Vorstellung der relevantesten Merkmale der besagten Ressource in ihrer aktuellen Version (v.1.0.6)²; (iii) Darstellung

² Bekanntlich befinden sich Wörterbücher und lexicografische Ressourcen aller Art in einem ständigen Entwicklungs- und Verbesserungsprozess, umso mehr, wenn es sich um ein Online-Informationssystem handelt. Da sich auch CombiDigiLex noch im Aufbau befindet, wurde die Ressource bis jetzt kaum genauer beschrieben. Mit der Publikation dieses Beitrages stellen wir zum ersten Mal die Ressource CombiDigiLex ausführlich vor und beschreiben die zahlreichen Herausforderungen, mit denen die Erstellung unserer lexicografischen Online-Ressource verbunden sind.



der methodologischen und technischen Schwierigkeiten, auf die wir nicht nur während der offiziellen Projektlaufzeit gestoßen sind, sondern die selbst zum Zeitpunkt der Abfassung dieses Artikels noch teilweise offene Fronten bilden.

Der Name des Projekts, den wir mit dem Akronym CombiDigiLex abkürzen, fasst bereits die wesentlichen Merkmale des Forschungsvorhabens zusammen: „Kombinatorik in lexikalisch-semantischen Paradigmen im Kontrast. Empirische Studien und Digitalisierung für den Fremdsprachenerwerb in germanisch-iberoromanischen Kontexten“. Das Hauptziel dieses Projekts, in dessen Rahmen die hier vorgestellte Ressource entwickelt wird, war die Durchführung von kontrastiven Studien einer Reihe von Phänomenen im Zusammenhang mit dem kombinatorischen Potenzial signifikanter Lexeme des Deutschen, Spanischen und Portugiesischen im Vergleich auf der Grundlage einer für lexikologische Studien geeigneten empirischen Methodik. Unsere Ergebnisse beruhen auf einer Auswahl von Vorkommen aus großen Referenzkorpora der drei für die Ressource ausgewählten Sprachen in vier Textsorten: Presse, Belletristik, wissenschaftliche Literatur und Gebrauchsliteratur. Für die erste Arbeitsphase haben wir uns vorerst auf einige wenige Verballexeme aus drei semantischen Verbklassen konzentriert: KOMMUNIKATION, EMOTION und TRANSFER.

Die zahlreichen Herausforderungen, mit denen wir uns im Rahmen des Projekts von Beginn an konfrontiert sahen und noch teilweise heute sehen, waren v.a. die Festlegung einer spezifischen korpusbasierten Methodik für die Analyse unter Berücksichtigung des mehrsprachigen und kontrastiven Charakters, die Koordinierung, die die Zusammenarbeit eines Teams von 12 ForscherInnen aus 4 Ländern und 8 verschiedenen Institutionen erforderte sowie die sprachtechnologische Modellierung und Erfassung der Daten und die effektive Übertragung der Ergebnisse auf eine Benutzeroberfläche mit einer dreisprachigen Schnittstelle und verschiedenen Suchoptionen³.

Aus einer Reihe von empirischen Studien zum Verhalten der NutzerInnen von Online-Wörterbüchern und lexikografischen Ressourcen und deren Anforderungen, insbesondere im Fall von Lernenden des Deutschen bzw. Spanischen als Fremdsprache (=DaF bzw. ELE) ab dem B1- Niveau konnten für die Erstellung unserer Ressource wichtige Erkenntnisse erlangt werden (Meliss, 2015; Fernández Méndez *et al.*, 2016; Mas Álvarez *et al.*, 2017). Die durch die Benutzerstudien u.a. auch aufgedeckten Unzulänglichkeiten der bestehenden lexikografischen Werke, die in Abschnitt 2 exemplarisch erläutert werden, einerseits und die aktuellen Anforderungen an die moderne Lernerlexikografie (Runte, 2015) andererseits haben uns dazu veranlasst, eine mehrsprachige lexikografische Ressource zu entwickeln, die das kombinatorische Potenzial lexikalischer Einheiten, so wie es sich in Korpusdaten manifestiert, in den Mittelpunkt des Informationsangebotes stellt. Der Ausgangspunkt des Artikels ist daher der Lernerkontext, in dem die Ressource CombiDigiLex entwickelt wird. Es folgt eine Darstellung der Ressource selbst, ihrer Struktur und ihrer Merkmale (Abschnitt 3). In Abschnitt 4 gehen wir deutlich auf die Herausforderungen und Schwierigkeiten ein,

³ Die erste Entwicklungsphase der Ressource ist noch nicht abgeschlossen und gerade bezüglich der Benutzeroberfläche sind noch Optimierungen notwendig.



mit denen wir bei der Planung und Erstellung der Ressource konfrontiert wurden, und abschließend ziehen wir einerseits ein Resümee und geben andererseits einen Ausblick mit einigen Desiderata für die zukünftige korpusbasierte Lernerlexikografie im multilingualen Kontext.

1. Grenzen der zweisprachigen Lernerlexikografie: Die Benutzerperspektive

Beim formalen Fremdsprachenerwerb wird die Verwendung von Wörterbüchern und lexikografischen Hilfsmitteln sowohl bei der Rezeption als auch bei der freien und angeleiteten Produktion, vor allem ab dem Niveau B1, gefördert, wobei mehrere Ziele verfolgt werden, die in Verbindung mit der Entwicklung einer digital-lexikografischen Kompetenz stehen: (i) die Lernenden sollen für die semantische Präzision und die genaue Verwendung von Lexik durch das Erlernen von Disambiguierungsstrategien sensibilisiert werden; (ii) die Nutzung von lexikografischen Ressourcen erleichtert sowohl die Aktivierung des neuen Lexikons als auch die Verwendung von Synonymen und das Bewusstsein für die sprachliche Variation auf unterschiedlichen Ebenen; (iii) die lexikografischen Konsultationen sollten Zweifel über Morphologie, Syntax und Sprachgebrauch beseitigen (Runte, 2015, S. 56f; Meliss, 2015, S. 423f; Meliss *et al.*, 2019, S. 140f).

Unterschiedliche Verwendungskontexte stellen unterschiedliche Anforderungen an lexikografische Werkzeuge. Im Bereich des Fremdsprachenerlernens sind einerseits die unterschiedlichen Anforderungen an die Situationen der Rezeption und Produktion und andererseits die spezifischen Anforderungen für das Hin- und Herübersetzen zu nennen. Doch allzu oft werden die Erwartungen der Nutzer an die zur Konsultation ausgewählten Nachschlagewerke nicht erfüllt und es lassen sich spezifische Probleme bei der Konsultation in der einen oder anderen Benutzungssituation aufdecken. Die L2-Lernenden stellen sich v.a. in fremdsprachigen Produktionskontexten, die hier im Mittelpunkt des Interesses stehen, immer wieder Fragen dieser Art: Welches Verb wähle ich aus dem Angebot aus? Wie verwende ich dieses Verb? Mit welchen anderen Elementen lässt sich dieses Verb kombinieren?

1.1. Recherche(probleme) in einem zweisprachigen Wörterbuch

Der Rechercheprozess in zweisprachigen Ressourcen steht sowohl in Verbindung mit einer Such- und Finde- als auch mit einer entsprechenden Auswahlhandlung (Meliss, 2014; González Ribao und Meliss, 2015). Die entscheidende Frage ist: Können die Lernenden das finden, was Sie suchen? Laut der Ergebnisse der zahlreichen empirischen Benutzungsstudien gehen wir davon aus, dass die Nutzung von zweisprachigen Online-Ressourcen eine prototypische Nutzungssituation widerspiegelt (Fernández Méndez *et al.*, 2016, S. 79-80; Meliss *et al.*, 2019, S. 135; Taborek, 2019, S. 212-213). Als Ausgangspunkt stellen wir uns daher mehrere Fragen: Warum liefert die Suche nicht immer gute Ergebnisse? Was fehlt in den vorhandenen Wörterbüchern? Was brauchen die Benutzer, um von einer lexikografischen Suche zu profitieren? Was sollte eine solche Ressource bieten?



Da das Verb im Deutschen eine hohe Anzahl von Lesarten aufweist, müssen zur besseren Disambiguierung die beiden Suchrichtungen Deutsch → Spanisch / Spanisch → Deutsch durchgeführt werden, um mittels Rückübersetzung (*escribir* → X) folgende Informationen zu erhalten:

- Mögliches Entsprechungsangebot in beiden Sprachrichtungen
- Syntaktisch-semantische Disambiguierung in der Zielsprache (L2=Spanisch)
- Überprüfung bestimmter syntaktischer Distributionsbeschränkungen in der Zielsprache

In der Richtung Spanisch → Deutsch (Abbildung 2) weist die besagte zweisprachige Ressource für spezifische Probleme in Produktionssituationen auch keine zufriedenstellende Information auf.

- unzureichende syntaktische und semantische Disambiguierung zu *escribir*
- Mangel an Information zu lexikalisch-semantischen Distributionsbeschränkungen
- Unzureichende illustrative Beispiele

Abbildung 2

Screenshot von *escribir* (Ausschnitt) und Entsprechungen im Deutschen in dem zweisprachigen PONS-online Wörterbuch und Kommentare zu dem Informationsangebot

The screenshot shows the dictionary entry for 'escribir' in the PONS online dictionary. The entry is divided into three sections: 1. *escribir* (apuntar), 2. *escribir* (redactar), and 3. *escribir* (comunicar). Each section lists Spanish phrases and their German equivalents. Two red boxes highlight specific information gaps:

- Mangel an Informationen zu der Bedeutung und Strukturmuster**: This box highlights the first section, where the German equivalent 'schreiben' is listed for 'escribir' without further context or structure information.
- Mangel an Informationen zu der lexikalisch-semantische Abdeckung des Objekts**: This box highlights the second section, where the German equivalent 'verfassen' is listed for 'escribir' without further context or semantic information.

Um die spezifischen Probleme und Bedürfnisse der Wörterbuchbenutzenden beider Ausgangssprachen genauer zu dokumentieren, erstellten wir eine Pilotumfrage für deutsch- und hispanophone L2-Lernende des Deutschen bzw. Spanischen (B1-Niveau), in der beide Sprachgruppen u.a. gebeten wurden, ein Foto einer Bibliotheksausleihe mit Hilfe

eines zweisprachigen Online-Wörterbuches detailliert zu beschreiben: „Was machen die Personen auf dem Bild?“. Ziel dieser Frage war, herauszufinden, wie die unterschiedlichen Lernenden dieselbe Situation beschreiben und welche Probleme sie trotz Verwendung eines zweisprachigen Wörterbuchs nicht lösen können. Auf diese Weise sollte durch die Verwendung eines Bildes für beide Sprachrichtungen eine onomasiologische Perspektive gefördert werden⁵. Exemplarisch kommentieren wir hier ausgewählte Beobachtungen bezüglich der Ergebnisse im ELE-Produktionskontext. Die 64 Spanischlernenden, von denen wir Antworten erhielten, produzierten einen hohen Prozentsatz an korrekten Satzstrukturen (72%), aber trotz der Verwendung des Wörterbuchs waren 28% der Äußerungen nicht akzeptabel, mit Strukturen des Typs (1-3):

- 1) *Una joven quiere *prestar un libro de la biblioteca*⁶
- 2) *Hay una chica que quiere *prestar un libro*
- 3) *Una mujer quiere *alquilar un libro en la librería*

Betrachtet man die Unterstützung durch das PONS-Wörterbuch (das in der Umfrage empfohlen wurde), so stellt man Folgendes fest: Die erste verfügbare Bedeutungsentsprechung ist diejenige, die offensichtlich zu einem hohen Prozentsatz von fehlerhaften Äußerungen geführt hat, da in dem Informationsangebot des Wörterbuches nicht berücksichtigt wurde, dass die in der Zielsprache geforderte syntaktische Struktur sich von der Ausgangssprache (Deutsch) unterscheidet (Abbildung 3).

Abbildung 3

Screenshot von *ausleihen* in dem zweisprachigen PONS-online Wörterbuch und Kommentare

Suchergebnisse

ausleihen (verleihen)	prestar	28% nehmen diese Bedeutung an
ausleihen (verleihen)	aviar <i>amer.</i>	
ausleihen (sich borgen)	tomar prestado	
jdm. etw. ausleihen	prestar algo a alguien	67 % der Antworten 2% der Antworten
dieses Buch habe ich mir von einem Freund ausgeliehen	este libro me lo ha prestado un amigo	
dieses Buch habe ich mir aus [o. in] der Bücherei ausgeliehen	he sacado este libro de la biblioteca	Nur 1%.

⁵ Die Ergebnisse werden ausführlich in Fernández Méndez (im Dr.) beschrieben.

⁶ Der Fehler in (1) und (2) besteht in der Verwendung eines Verbs, das es nicht zulässt, dass der Nutznießer der Handlung das grammatische Subjekt ist. Der Fehler in (2) besteht in der Verwendung des falschen Verbs *alquilar*, das in einigen Fällen dem deutschen Verb *ausleihen* entsprechen kann, aber hier den deutschen Verben *mieten/vermieten* entsprechen würde.

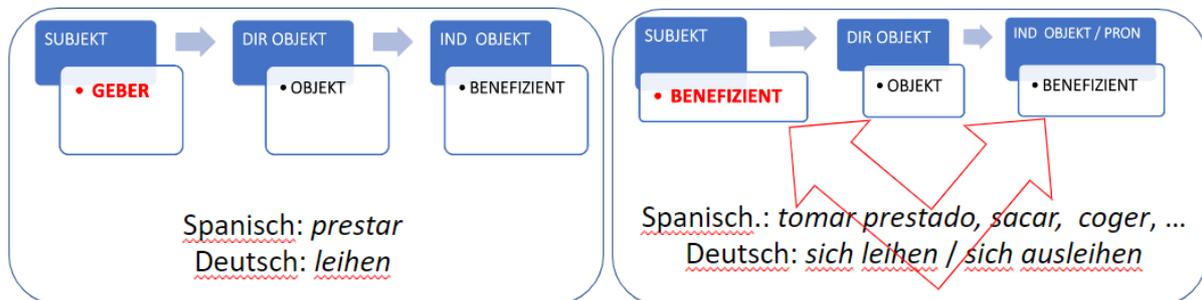
1.2. Welche Lücke versucht CombiDigiLex zu füllen?

Sowohl aus unseren Umfrageergebnissen, die im Einklang mit vielen anderen Ergebnissen von empirischen Benutzerstudien des letzten Jahrzehnts stehen, sowie den dargestellten konkreten Rechercheproblemen in L2-Produktionssituationen (vgl. 2.1) leiten sich Desiderata für die Erstellung einer neuen Ressource ab. Die Erstellung von CombiDigiLex knüpft genau hier an, indem das Ziel verfolgt wird, konkret für L2-Produktionssituationen im multilingualen Kontext neue Wege anzubieten. Die Ressource soll v.a. bei der semantisch-syntaktischen Disambiguierung der Verwendungsmöglichkeiten helfen. Im Zentrum stehen daher die Informationen zum kombinatorischen Potenzial, welche durch die Argumentstrukturen (Anzahl und Art der Argumente) beschrieben werden. Am Beispiel von (*sich*) (*aus*)*leihen* in Verbindung mit dem Konzept LEIHEN soll dies erklärt werden. Dieselbe bzw. ähnliche dreiwertige syntaktische Verbalstruktur kann unterschiedliche Argumentstrukturen (AST_A und AST_B) aufweisen und damit zu verschiedenen Lesarten und dementsprechend zu unterschiedlichen Entsprechungen im Spanischen führen (vgl. Beispiele 4 und 5 sowie Abbildung 4).

- 4) *Kannst du mir bitte dein Fahrrad leihen?* → „prestar“
 AST_A: Subjekt = GEBER
- 5) *Außerdem leiht sie sich häufig Geld von diesem und jenem und zahlt nur selten ihre Rechnungen.* → „tomar prestado“, „sacar / coger(de la biblioteca)“
 AST_B: Subjekt = BENEFIZIENT

Abbildung 4

Argumentstrukturen A (links) und B (rechts) und ihre Auswirkungen in beiden Sprachen am Beispiel des LEIHENS



Wir stellen uns daher vor, dass unsere Ressource im Abschnitt KOMBINATORIK einen Hinweis auf die Grenzen der Verwendung und im Abschnitt VERGLEICH einen kontrastiven Kommentar enthält (vgl. Abschnitte 3.2. und 3.3), der verdeutlicht, warum z.B. sprachlichen Äußerungen wie *„presto un libro de la biblioteca“ in dem Paradigma TRANSFERENZ oder *„escribir un examen“ in dem Paradigma KOMMUNIKATION (vgl. Abschnitt 2.2.) nicht gültig sind.

Die Nützlichkeit von ein-, zwei- oder mehrsprachigen Online-Ressourcen, die über einen kostenlosen Zugang lexikalische Daten aus einer semasiologischen Perspektive anbieten und die u.a. auch in der Lage sind, (automatisch)

Beispiele für Parallelübersetzungen anzubieten, wird immer deutlicher und konnte in den vorausgehenden Abschnitten beispielhaft aufgezeigt werden. Obwohl die aus solchen Ressourcen gewonnenen Informationen für die Produktionsaufgaben beim Fremdsprachenerlernen von großem Interesse sind, stellen sie für eine optimale Nutzung einige Probleme dar. Erstens erfordern sie im Allgemeinen, dass die Lernenden in der Lage sind, unterschiedliche Lesarten zu unterscheiden, eine Aufgabe, zu der sie oft nicht fähig sind. Da die Suche meistens semasiologisch ausgerichtet ist, sind mehrere Versuche erforderlich, um die gewünschte ideale Verwendung eines Lexems in einer bestimmten Kombination zu finden. Bei zwei- und mehrsprachigen Ressourcen stehen die L2-Lernenden außerdem vor dem Problem, aus verschiedenen Entsprechungsangeboten eine adäquate auswählen zu müssen. Außerdem sind die entsprechenden automatisch selektierten Übersetzungsbeispiele, die meistens aus Parallelkorpora stammen, oft unzureichend, entweder weil sie aus schlecht gepflegten Film- oder Website-Untertiteln stammen oder weil sie aus thematisch begrenzten Textsorten entnommen sind, wie z.B. denen von offiziellen Dokumenten. Schließlich obliegt es dem Benutzer, aus einer Liste von Beispielen, die nicht nach linguistischen Kriterien geordnet ist, die relevanten Elemente der lexikalisch-grammatischen Kombinationsmöglichkeiten abzuleiten.

Der Vorteil, schnell an viele Beispiele zu kommen, ist direkt verbunden mit den Nachteilen, die die automatische Extraktion von Informationen mit sich bringt. Die redaktionelle Arbeit von Fachleuten, die durch linguistische Analysen Ordnung und Struktur in lexikalische Phänomene an der Semantik-Syntax-Schnittstelle liefern, fehlt. Der Zugang zum kombinatorischen Profil der lexikalischen Einheiten in verschiedenen Sprachen ist derzeit nur durch manuelle Auswahl und Analyse durch Spezialisten möglich. Die Ressource CombiDigiLex stellt sich diesen Anforderungen und Herausforderungen.

2. Die lexikografische Ressource CombiDigiLex

Im Folgenden wird CombiDigiLex, die als offene, multilinguale Ressource konzipiert ist, vorgestellt⁷. Es handelt sich um ein über das Internet frei zugängliches, lexiko-grammatisches Informationssystem, welches durch einen projektinternen kollaborativen Arbeitsprozess aufgebaut und ständig erweitert wird⁸. In der aktuellen Version (Prototyp v.1.0.6), auf die wir uns hier genauer beziehen, spiegelt die Ressource den Stand der ersten Erstellungsphase wider. Im Mittelpunkt der folgenden Ausführungen steht die komplexe lexikografische Struktur⁹.

⁷ Zugriff auf die Ressource über: <http://combidigilex.usc.gal/index.php#> [letzter Abruf: 27.09.2021].

⁸ Zu den unterschiedlichen lexikografischen Aspekten verweisen wir u.a. auf die Sammelbände von Granger und Paquot, 2012; Gouws *et al.*, 2013; Klosa und Müller-Spitzer, 2016 und Fuertes Olivera, 2018.

⁹ Da wir im Rahmen dieses Beitrags weder auf die theoretischen Grundlagen noch auf die Methodik eingehen können, die wir zur Gewinnung der Daten unserer qualitativen und quantitativen Analysen angewandt haben, laden wir alle Interessierten ein, die detaillierten Informationen in dem von uns herausgegebenen Sammelband zu konsultieren, der verschiedene Beiträge zu diesem Projekt enthält (Meliss *et al.*, im Dr.).

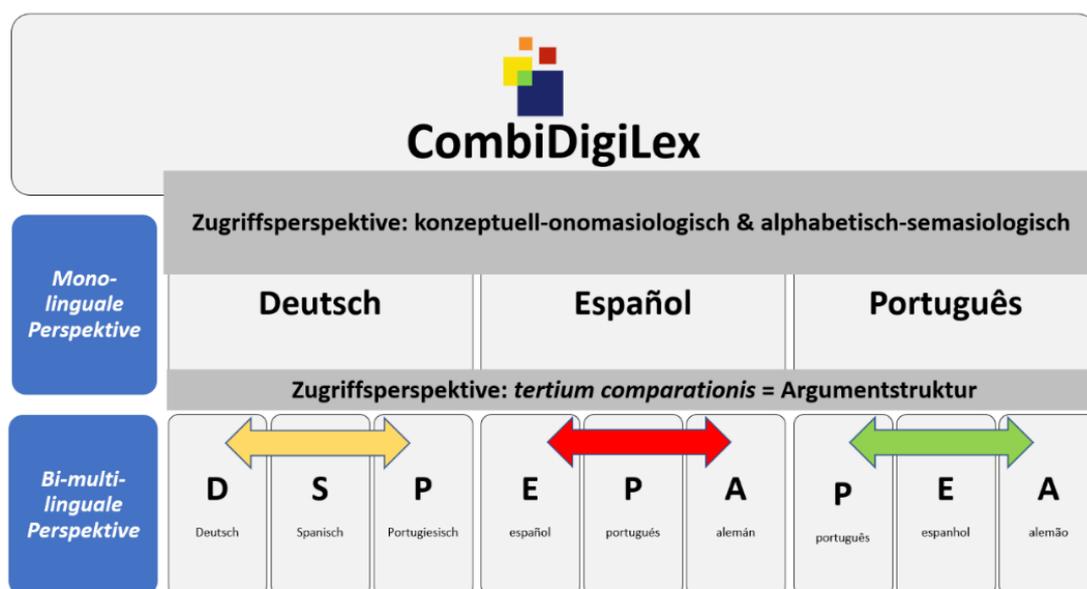


2.1. Die Makrostruktur: Onomasiologie vs. Semasiologie

Die Makrostruktur von CombiDigiLex ist durch den benutzeradaptierten Zugriff über die drei z.Z. behandelten Sprachen der Ressource relativ komplex. Die Auswahl einer der drei Sprachen bedeutet allerdings nicht, dass das Informationsangebot in allen drei Sprachen identisch sei. Vielmehr steht hinter jeder Sprache das jeweilige einzelsprachige lexikalische Angebot, d.h. die Liste der behandelten lexikalischen Einheiten in Verbindung mit ihren lexikalischen Merkmalen und den empirisch erhobenen Daten. Die Auswahl der einen oder anderen Sprache impliziert damit zunächst nur ein monolinguales Informationsangebot, welches erst in einem spezifischen Modul zum VERGLEICH mit den Kontrastsprachen in Relation tritt. In diesem Sinne könnte man die Ressource in ihrer Gesamtheit auch als lexikografisches Portal verstehen, welches drei monolinguale lexikografische Ressourcen umfasst, die über eine gemeinsame Oberfläche die Such- und Auswahloptionen vereint. Darüber hinaus wird über ein Navigationsangebot zum VERGLEICH - der über die entsprechenden Argumentstrukturen als *tertium comparationis* erfolgt - die Querverbindungen zu den anderen Sprachen hergestellt, um Konvergenzen und Divergenzen aufzuzeigen (Abbildung 5).

Abbildung 5

Makrostruktur der Ressource CombiDigiLex

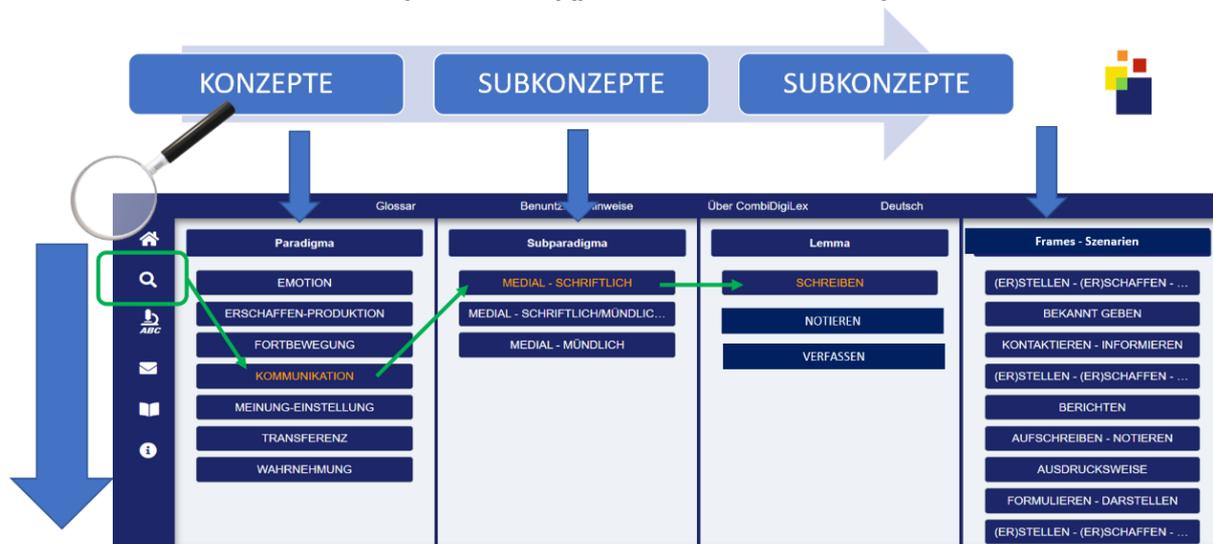


2.1.1. Eine der relevantesten Merkmale, die CombiDigiLex von anderen lexikografischen Ressourcen unterscheidet, ist der konzeptuell-onomasiologische Zugriff auf die Information: ausgehend von einer Idee bzw. einem Konzept sollen die verschiedenen Lexikalisierungsmöglichkeiten aufgezeigt werden. Diese Perspektive ermöglicht v.a. den Informationszugriff für L2-Produktionssituationen, in denen zunächst ein passendes Lexem aus dem Angebot

ausgewählt werden muss. Das Informationsangebot für diese Art des Zugriffs erfolgt sowohl über eine horizontale als auch über eine vertikale Achse. Die Auswahl erfolgt von links nach rechts vom Allgemeinen zum Spezifischen. Die zunehmende Spezifizierung soll den konkreten Kommunikationsbedürfnissen der Benutzenden entgegenkommen (Abbildung 6).

Abbildung 6

Onomasiologisch orientierte Zugriffsstruktur mit zwei Achsen in CombiDigiLex.



Die **horizontale Achse** bietet für ein (Sub)Paradigma, welches über konzeptuelle Einheiten definiert wird, ein Angebot von Lemmata, die in direkter semantischer Verbindung stehen, an. So können z.B. die deutschen Lemmata *schreiben*, *notieren* und *verfassen* dem Paradigma KOMMUNIKATION und dem Subparadigma MEDIAL SCHRIFTLICH zugeordnet werden. Ausgehend von einem konkreten Lemma werden die verschiedenen Frames bzw. Szenarien zusammen mit den dazugehörigen Konzepten bzw. konzeptuellen Einheiten zugeordnet. Die Auswahl des einen oder anderen Frames ermöglicht den Zugriff auf eine der möglichen Lesarten des ausgewählten Lemmas. Zur Illustration dient hier das Frame: ERSTELLEN - ERSCHAFFEN dem die Lesart 1 von *schreiben* im Sinne von „einen Text schriftlich produzieren, erschaffen“ zugeordnet wird. Über die **vertikale Achse** erfolgt der Zugriff auf unterschiedliche Listen von (Sub)paradigmen sowie auf die dazugehörigen Lemmata und unterschiedlichen Frames.

2.1.2. Zusätzlich zu dieser konzeptuell-onomasiologischen Perspektive steht auch ein alphabetisch-semasiologischer Zugriff zur Verfügung. Bei der Auswahl eines der aufgelisteten Lemmata wird zunächst eine allgemeine lesartenübergreifende Überblicksinformation angeboten, die verschiedene lexikalische Bereiche und lexikografische Angabeklassen umfasst. Obwohl diese Information schon Bestandteil der lexikografischen Mikrostruktur ist, soll sie an dieser Stelle an dem Eintrag zu *schreiben* illustriert werden (Abbildung 7): morphosyntaktische Basisinformation zusammen mit einem allgemeinen Kommentar (*ad 1*), quantitative Information bezüglich der Beleganzahl verteilt auf

die analysierten Textsorten und *Genus Verbi* (ad 2) und ein Navigationsangebot zu dem spezifischen Informationsangebot jeder einzelnen Lesart eines Lemmas über die Auswahl eines der Szenarien und (Sub)Konzepte (ad 3).

Abbildung 7
Semasiologisch orientierte Zugriffsstruktur von CombiDigiLex: Illustration am Beispiel von schreiben

2.2. Die Mikrostruktur: Einzelsprachige lexikalische Information: Bedeutung und Kombinatorik

Der Zugriff zu der einzelsprachig relevanten, lesartenspezifischen lexikalischen Information erfolgt sowohl aus einer onomasiologisch als auch einer semasiologisch orientierten Perspektive jeweils über entsprechende Frames und konzeptuelle Einheiten (Abbildungen 8 und 9).

2.2.1. Die lexikografische Mikrostruktur jeder Lesart besteht aus einer Menge von lexiko-grammatischen Informationsebenen, die im Folgenden kurz am Beispiel von *schreiben 2* im Sinne von „etwas schriftlich bekannt geben“ präsentiert werden (Abbildung 8).

Abbildung 8
Die Mikrostruktur von CombiDigiLex am Beispiel von schreiben 2

Eine graphische Darstellung zeigt die Summe aller konzeptuellen Einheiten, die in Verbindung mit der Lesart stehen (ad 1). Dieser Graph befindet sich noch in einer Aufbauphase und soll mit interaktiven Funktionalitäten versehen werden. Allgemeine grammatische Informationen (ad 2) bestehend aus: einer Bedeutungsdefinition, einer semantischen Etikettierung, die in den meisten Fällen durch Synonyme erfolgt, und quantitativer Daten, die das Auftreten von *Genus Verbi* und den unterschiedlichen Textsorten in Korrelation setzen. Die Korpusbelege und die entsprechenden Quellenangaben können jeweils über die quantitativen Daten angesteuert werden (ad 3). Der Export ist möglich über entsprechende Funktionalitäten (Abbildung 9). Ausführliche Information zu dem Kombinationspotenzial ist über den Schaltknopf KOMBINATORIK möglich (ad 4) (vgl. hierzu genauere Ausführungen in Abschnitt 3.2.2.). Auf die kontrastive Information kann über den Schaltknopf VERGLEICH zugegriffen werden (ad 5) (vgl. hierzu ausführliche Information in Abschnitt 3.3.).

Abbildung 9

Ausschnitt aus CombiDigiLex: Korpusbelege zu schreiben 2

The screenshot displays the CombiDigiLex interface for the lemma 'schreiben 2'. The main content area is divided into several sections:

- Glossar:** 'schreiben 2 <=> BEKANNT GEBEN' with a definition: 'Bedeutungsdefinition: Jemand (Autor, Schreiber, ...) gibt etwas (Information, Nachrichten ...) schriftlich bekannt.' and a comment: 'Kommentar: Diese Lesart konnte in knapp einem Drittel aller analysierten Belege dokumentiert werden. Sie ist fast ausschließlich im Aktiv belegt. In Passivkonstruktionen konnte diese Lesart sogar in mehr als einem Drittel der Belege identifiziert werden, während narrative Texte nur ein Viertel der Belege dieser Lesart aufweisen konnten.'
- Benutzungshinweise:** A table showing the distribution of 'schreiben 2' across different text types (Texttyp) and grammatical forms (Genus Verbi).
- Table:**

schreiben 2	Texttyp			
Genus Verbi	P	BE	WI	GL
Aktiv	36	21	66	25
Passiv	0	0	2	1
andere Passivformen	0	0	0	0
Aktiv unpersönlich	0	0	0	0
abs. Frequenz	36	21	68	27
- Über CombiDigiLex:** 'SCHREIBEN <=> Bekannt geben' and 'Gebrauchsliteratur <=> Activa'.
- Beispiel:** A list of four corpus examples with their respective sources (Quelle).

2.2.2. Die ausführliche Darstellung des **Kombinationspotenzials** jedes Lemmas in Verbindung mit jeder einzelnen Lesart ist eins der Hauptanliegen der Ressource CombiDigiLex. Über den Schaltknopf KOMBINATORIK wird in unterschiedlichen Modulen diesbezüglich Information angeboten (Abbildung 10):

Abbildung 10
Module zur Kombinatorik in CombiDigiLex am Beispiel von schreiben 2

The screenshot displays the CombiDigiLex interface for the verb 'schreiben'. On the left, 'Modul 1' (circled in green) provides semantic information, including a definition and examples. On the right, three modules (2, 3, and 4, also circled in green) show argument structures for different text types: 'Allgemeine Argumentstruktur', 'Kanonische Argumentstruktur', and 'Argumentstruktur mit Variation'. Each module displays abstract formulas and lists of semantic roles or arguments.

Das **Modul 1** auf der linken Seite fasst die semantische Information einer Lesart zusammen und bietet eine Auswahl von Muster- bzw. Kompetenzbeispielen an. Diese sind ausgewählte Korpusbelege, die zur Illustration und besserem Verständnis redaktionell bearbeitet bzw. vereinfacht wurden. Dieses Modul stellt die Schnittstelle zwischen der Information zur Semantik und Syntax dar.

Die Module (2) bis (4) auf der rechten Seite bieten feingranulare qualitative und quantitative Information zu den Argumentstrukturen, den entsprechenden Satzmustern und den morphosyntaktischen und lexikalischen Besonderheiten jedes Aktanten in Korrelation zu ihrem Auftreten in den einzelnen Textsorten an. Auch hier ist in allen Fällen der Zugriff auf die Korpusbelege über die tabellarische Darstellung ihrer quantitativen Verteilung auf die einzelnen Textsorten möglich.

So stellt das **Modul 2** folgende abstrahierte Information zu der allgemeinen Argumentstruktur in Verbindung mit den vier analysierten Textsorten zur Verfügung:

- Eine abstrakte Argumentstrukturformel weist die maximale Anzahl und Art der möglichen semantischen Rollen/Argumente in Verbindung mit der entsprechenden Lesart, die einem konkreten Szenarium zugeordnet wurde, auf.
- Außerdem werden die einzelnen semantischen Rollen gesondert aufgeführt. Da sie entsprechend ausgezeichnet werden, soll auch eine sich in Planung befindliche weiterführende Suchfunktion, in Zukunft entsprechende Ergebnisse in Verbindung mit den einzelnen Rollen anbieten.
- Über eine Strukturmusterformel werden die möglichen Aktanten eines Szenarios syntaktisch und semantisch definiert.
- Ein allgemeiner Kommentar dient für Hinweise bezüglich des Auftretens der einzelnen Aktanten.

Das **Modul 3** verweist auf konkret realisierte und belegte Argumentstrukturmuster. Einerseits werden die **kanonischen Argumentstrukturmuster** sowohl durch eine abstrakte Argumentstrukturformel als auch durch eine Strukturformel aufgezeigt. Es handelt sich um die aktiven Muster, die am häufigsten belegt werden konnten und die aus den Basisrollen, die für die jeweilige Szene relevant sind, bestehen. Durch zwei Kommentartypen werden die Informationen systematisiert: ein allgemeiner Kommentar fasst die wichtigsten argumentstrukturellen Besonderheiten zusammen und ein weiterer morphosyntaktischer Kommentar weist auf Besonderheiten bezüglich der Realisierungsmöglichkeiten der einzelnen Aktanten hin. Hier sind u.a. Informationen zu Präpositionen und dem regierenden Kasus oder satzförmige Realisierungsmöglichkeiten von Interesse.

Das **Modul 4** listet die Argumentstrukturmuster, die eine gewisse **Variation** besitzen, auf. Die Variation bezieht sich einerseits auf alle nicht aktivisch auftretenden Realisierungsmöglichkeiten und andererseits auf die Beteiligung von Rollen, die nur peripher zu der entsprechenden Szene zugeordnet werden können. Diese sind z.B. bei der Szene KOMMUNIKATION – MEDIAL SCHRIFTLICH – PRODUKTION die Rollen INSTRUMENT (I) und ART UND WEISE (MO). In diesen Fällen wird auf die Angabe einer Strukturformel verzichtet, aber besonderen Wert auf die zwei oben erwähnten Kommentartypen gelegt.

2.3. Mediostruktur: Von der monolingualen zur multilingualen und kontrastiven Perspektive

Durch die verschiedenen internen Interrelationen zwischen den einzelnen Modulen und Informationsebenen, die die lexikografische Mediostruktur ausmacht, wird auch eine Recherche zu der vergleichenden Information ermöglicht, die die einzelnen Sprachen, die in der Ressource behandelt werden, in Relation setzt. Dieser Vergleich ist möglich über ein *tertium comparationis*, welches sich aus den Argumentstrukturen ergibt. In dem Fall der Szene KOMMUNIKATION – MEDIAL SCHRIFTLICH – PRODUKTION wird eine Argumentstruktur mit den folgenden zwei Rollen angenommen: SCHREIBER (ESCRITOR) + INHALT (CONTENIDO). Über diese Struktur ist eine Verbindung zwischen *schreiben 2* (Lesart, die diesem Szenario zugeordnet werden kann) zu Entsprechungen in anderen Sprachen, wie z.B. zu *escribir 2* im Spanischen mit derselben abstrakten Argumentstruktur **S_I** möglich, ganz unabhängig von den syntaktisch-funktionalen Merkmalen der einzelnen Aktanten (Abbildung 11).



Abbildung 11
Sprachvergleich in CombiDigiLex

The screenshot displays the CombiDigiLex interface. At the top, there is a navigation bar with 'Glossar', 'Benutzungshinweise', and 'Über CombiDigiLex'. Below this, a sidebar on the left contains icons for home, search, and other functions. The main content area is divided into two sections. The top section shows the German entry for 'schreiben 2 \leftrightarrow BEKANNT GEBEN', including its definition, usage instructions, and a comment. A green box highlights the 'Vergleich' button in the top right of this section. The bottom section, titled 'Vergleichsinformation \leftrightarrow SCHREIBEN \leftrightarrow BEKANNT GEBEN', is a table comparing the German and Spanish entries. The table has columns for 'Deutsch', 'Spanisch', and 'Portugiesisch'. The German and Spanish columns are highlighted with a green box. The table contains information such as 'Abstrakte Formel', 'Semantische Rollen', 'Strukturmuster', and 'Kommentar'. Green arrows point from the 'Vergleich' button to the comparison table.

2.4. Weitere Informationsmodule

Folgende weitere Informationsmodule, die sich als lexikografische Außen- oder Umtexte bezeichnen lassen, befinden sich z.Z. noch im Aufbau und werden im oberen horizontalen Menü in der jeweils selektierten Sprache angeboten:

- Ein Glossar, welches in den involvierten Sprachen der Ressource die relevanten grammatischen Termini und entsprechende Kurzerklärungen anbietet bzw. tabellarisch gegenüberstellt.
- Benutzungshinweise, die an der Benutzungssituation: L2-Produktion orientiert sind
- Projektrelevante, theoretische und methodologische Grundlageninformation

Dieses zusätzliche Informationsangebot spricht sowohl die BenutzerInnen in L2-Produktionssituationen als auch die wissenschaftliche *Community* an und soll einerseits als Hilfestellung für die Benutzung der Ressource CombiDigiLex und andererseits als Darstellung des Forschungsvorhabens dienen.

3. Herausforderungen

Im Folgenden sollen sowohl aus einer mono- als auch einer multilingualen Perspektive die unterschiedlichen Herausforderungen beschrieben werden, die mit der Erstellung einer korpusbasierten digitalen Ressource, welche auf manuellen Analysen und einem redaktionell bearbeiteten Informationsangebot beruht, einhergehen. Bezüglich der unterschiedlichen Phasen des komplexen lexikografischen Prozesses werden hier neben einigen methodologischen Herausforderungen (vgl. Abschnitt 4.1), technische/technologische Herausforderungen in Verbindung mit der

Datenmodellierung (vgl. Abschnitt 4.2) und Herausforderungen in Verbindung mit der benutzerorientierten Erstellung der Ressource an sich (vgl. Abschnitt 4.3) genannt, sowie auf Lösungsmöglichkeiten hingewiesen.

3.1. Methodologische Herausforderungen

Für die Zwecke von CombiDigiLex konnte weder auf bereits vorhandene aktuelle lexikografische und grammatikografische Ressourcen zur Darstellung der syntagmatischen Kombinatorik noch auf etablierte Methoden für kontrastiv angelegte, korpusbasierte Studien von Argumentstrukturen zurückgegriffen werden (vgl. González Ribao und Meliss, im Dr.). Sowohl die theoretischen Grundlagen als auch die anwendungsorientierten Ziele des Forschungsprojekts CombiDigiLex bilden den Hintergrund für die erforderliche Entwicklung einer korpusbasierten Methodik für lexikalische Fragestellungen an der Schnittstelle zwischen Semantik und Syntax im multilingualen Sprachvergleich. Die methodologischen Herausforderungen umfassen daher verschiedene Aspekte in Zusammenhang mit (i) der Korpusauswahl, der Datenbeschaffung und Erstellung einer empirischen Datengrundlage, (ii) der Bestimmung des Beschreibungsgegenstandes, (iii) der Erstellung eines speziell auf die Zwecke der Studien zugeschnittenen, korpusbasierten Daten- und Analyseinventars (iv) sowie der projektinternen kollaborativen Analysemethodik der Datenstrukturierung und dem multilingualen Vergleich und schließlich (v) dem *workflow*¹⁰.

(i) Die **Korpusauswahl** und die Erstellung einer empirischen Datengrundlage von möglichst vergleichbaren **Korpusstichproben** für kontrastive Studien zur Textgenrespezifität von Argumentstrukturen im Vergleich Deutsch-Spanisch/Portugiesisch steht in Zusammenhang mit der Thematisierung der Vergleichbarkeit der für die Studien herangezogenen Korpora und den damit verbundenen Anforderungen an das Prinzip der Vergleichbarkeit. Da das Identifizieren von Konvergenzen und Divergenzen im Sprachvergleich nur über eine vergleichbare empirische Datengrundlage erfolgen kann, war es erforderlich, sowohl spezifische Kriterien zur Auswahl der unterschiedlichen Korpora der zu analysierenden Sprachen als auch zur Sampleziehung aufzustellen (vgl. González Ribao, 2015; Trawiński und Kupietz, 2021; González Ribao *et al.*, im Dr.).

(ii) Die Bestimmung des **Beschreibungsgegenstandes** der Ressource durch die Auswahl von bestimmten Verbklassen und die Erstellung einer entsprechenden Stichwortliste wurde hauptsächlich aus den Anforderungen der L2-Perspektive abgeleitet. Aus den zahlreichen Verbklassen wurden zunächst folgende sieben Großklassen und daraus nur einige typische lexikalische Vertreter – vorzugsweise Einwortlemmata mit einer Basisstruktur ausgewählt:

- Fortbewegungsverben (*gehen, fahren, fliegen* etc.)
- Kommunikationsverben (*sagen, schreiben, fragen* etc.)¹¹

¹⁰ Zu einer ausführlichen Darstellung dieser Aspekte verweisen wir auf González Ribao *et al.* (im Dr.).

¹¹ Vgl. auch die kontrastiven Studien im Bereich der Kommunikationsverben von González Ribao und Proost (2015) und González Ribao (2021).



- Kreations-/Produktionsverben (*schreiben, produzieren, schaffen, herstellen, erstellen, kreieren* etc.)
- Meinungs-/Einstellungsverben (*meinen, denken, glauben, finden* etc.)
- Psychverben/Emotionsverben ((sich) *freuen, (sich) ängstigen, gefallen, hassen, lieben, mögen, (sich) fühlen* etc.)
- Transferverben (*geben, nehmen, leihen, mieten, schenken, borgen* etc.)
- Wahrnehmungsverben (*sehen, hören, fühlen, riechen, schmecken* etc.)

Für die erste Arbeitsphase war es erforderlich, zunächst eine reduzierte Anzahl von Zielverben für die Extraktion der Korpusbelege und ihre anschließende qualitative und quantitative Analyse zu selektieren¹². Die verwendeten Selektionskriterien stehen u.a. in Verbindung mit dem hohen Grad an Polysemie der betreffenden Verben, der erwartbar hohen Divergenz im Sprachvergleich und/oder der Relevanz ihrer Lesarten und Argumentstrukturen aus der Sicht der Fremdsprachendidaktik. Eine ausführliche Begründung der Verbauswahl kann in González Ribao *et al.* (im Dr.) nachgelesen werden.

(iii) Die Entwicklung eines geeigneten, möglichst umfassenden übereinzelsprachlichen **Annotationsschemas** mit morphologischen, morphosyntaktischen, syntaktischen und semantischen Parametern sowie die Erstellung der aus den Annotationen gewonnenen **Verbprofilen** zur Datenstrukturierung bilden den theoretischen Rahmen für die einzelsprachlichen Datenanalysen sowie den multilingualen Vergleich über abstrahierte Argumentstrukturen als *tertium comparationis*. Da das Projekt anstrebt, mittels korpusbasierter feingranularer Analysen Konvergenzen und Divergenzen im multilingualen Sprachvergleich zu ermitteln, die verschiedene sprachliche Ebenen umfassen, war die Entwicklung einer projektspezifisch zugeschnittenen Methode zur Ermittlung von quantitativen und qualitativen Daten in Verbindung mit dem verbalen Kombinationspotenzial für die Projektarbeit von COMBIDIGLEX unerlässlich. Vorbilder wie u.a. *FrameNet*¹³ (Fillmore *et al.*, 2003) wurden bei diesem Entwicklungsprozess zielorientiert genutzt und adaptiert. Außerdem konnte zusätzlich zu der lesartenspezifischen Kodierung von syntaktischen Strukturen sowohl die lexikalische Realisierung als auch die semantische Rolle der Verbargumente und in manchen Studien auch Merkmale des Verbalkomplexes textsortenspezifisch erfasst werden. Die Häufigkeit von unterschiedlichen Argumentstrukturmustern in Verbindung mit möglichen textsortenspezifischen Phänomenen im Sprachvergleich oder die morphosyntaktische Realisierung der einzelnen Argumente in Korrelation mit ihrer syntaktischen Funktion stehen hier nur exemplarisch für die verschiedenen empirischen Analysemöglichkeiten. Diese sind aus einer inter- und intralingualen Perspektive nur auf der Basis einer vergleichbaren Datengrundlage erfassbar.

¹² Die analysierten Lemmata der ersten Arbeitsphase sind hier in Fettdruck markiert. Die Ergebnisse der detaillierten Einzelverbstudien werden in dem Sammelband von Meliss *et al.* (im Dr.) vorgelegt: *schreiben/escibir* (Franco Barros und Meliss, im Dr.), *fragen/preguntar* (Egido Vicente, im Dr.), *fühlen/sentir/sentir* im Deutsch-Spanisch-Portugiesischen Vergleich (Mas Álvarez und Santos Palmou, im Dr.) und zu einer Auswahl von Transferverben im Deutsch-Spanischen Vergleich (Fernández Méndez, im Dr.).

¹³ Zu dem Nutzen von semantischen Frames im multilingualen Sprachvergleich siehe u.a. Boas (2009).



Für den multilingualen Kontext des Projekts hat sich diesbezüglich die Entwicklung einer übereinzelsprachlichen Metasprache als besonders komplex erwiesen. Eine geeignete Metasprache, die nicht nur für die linguistischen Annotationen, sondern auch für die Recherchemöglichkeiten auf das Informationsangebot der Ressource von Relevanz ist, sollte in einer benutzerorientierten Lerner-Ressource bestimmte Anforderungen erfüllen, denen wir nur schwerlich gerecht werden konnten.

(iv) Der gesamte *workflow* bzw. methodologisch-analytische Arbeitsprozess umfasst eine Reihe von unterschiedlichen Arbeitsschritten, die jeweils von bestimmten Mitgliedern des Projektteams durchgeführt werden. In Bezug auf die Datenbeschaffung und die händische Datenanalyse erweisen sich folgende, meist kollaborativ ausgeprägte Arbeitsschritte als besonders arbeitsintensiv und zeitaufwändig¹⁴:

- Extraktion der Korpusbelege
- Händische Annotation der einzelnen extrahierten Korpusbelege für jedes der ausgewählten Sprachen nach den verabredeten Kodierkonventionen und Erfassung der Daten in der dafür vorbereiteten Excelkodiertabelle. Dieser Arbeitsschritt erfolgt idealerweise immer von zwei Projektteammitgliedern parallel für die ausgewählten korrespondierenden Verben der unterschiedlichen Sprachen
- Anschließend gemeinsamer händischer Abgleich der Daten, um mögliche Diskrepanzen zu beheben
- Händische Datenstrukturierung und Erstellung von Wortprofilen durch unterschiedliche Abstraktionsprozesse
- Einzelsprachliche qualitative und quantitative Analysen der Wortprofile
- Qualitative Analysen im Sprachvergleich
- Aufbereitung der Daten für wissenschaftliche Zwecke (Publikation etc.)
- Aufbereitung der Daten zur Einbindung in die Ressource CombiDigiLex (vgl. Abschnitt 4.2)

3.2. Technische/technologische Herausforderungen

Schon im Projektantrag hatten wir die Notwendigkeit formuliert, für die Computerisierung bzw. die technische Umsetzung externe Unterstützung von Experten für Sprachtechnologien und natürliche Sprachverarbeitung zu beauftragen, da diese Bereiche die Fachkompetenz des Projektteams von CombiDigiLex überschreiten. Im Prozess der redaktionellen Wörterbucharbeit stellte sich die Einrichtung des lexikografischen Arbeitsplatzes bzw. Arbeitsumgebung für die technischen und lexikografischen Aspekte sowie die Modellierung der Daten als besonders komplex heraus. Die technologischen Arbeiten verfolgten drei Ziele:

- Die Entwicklung einer Referenzdatenbank für unsere Ressource, d.h. die Vorbereitung eines Systems, auf das alle Mitglieder des Teams von verschiedenen geografischen Standorten aus zugreifen können, ohne dass eine übermäßig detaillierte Vorbereitung erforderlich ist.
- Die Verbindung zwischen der Datenbank und unseren Annotationstabellen zur Analyse der extrahierten Korpusbelege.
- Die Gestaltung des endgültigen Online-Konsultationsprodukts mit einer geeigneten Benutzeroberfläche

¹⁴ So liegen z.B. für das Verpaar *schreiben/ escribir* im deutsch-spanischen Vergleich eine Summe von 800 Korpusbelegen zur händischen Analyse vor: 100 pro Textsorte (Presse, Belletristik, Gebrauchsliteratur und Wissenschaftstexte) und Sprache.

Der erste mühsame Prozess bestand darin, die beauftragten Sprachtechnologieexperten dazu zu bringen, unsere Bedürfnisse zu verstehen, da wir nicht vorhatten, ein traditionelles Wörterbuch zu erstellen, sondern eine lexikografische Ressource mit drei beteiligten Sprachen, verschiedenen Ebenen der Informationstiefe und einem gemischten Repertoire an Angabeklassen, das eine unterschiedliche Kategorisierung der Daten auf der lexikalischen Ebene an der Syntax-Semantik-Schnittstelle erlaubt. Wie wir im vorangegangenen Abschnitt gesehen haben, liegt unser Interesse in der konzeptuell-onomasiologischen Richtung, was ein zusätzliches Hindernis bei der Anpassung der gewählten Software, TshwaneLex (kurz TLex), darstellte. TLex ist ein Paket professioneller Softwareanwendungen für die Erstellung von Wörterbüchern oder Terminologielisten. Wie auf der Website erklärt wird, haben es viele Verlage als Hilfsmittel eingesetzt: „TshwaneLex ist eine professionelle, langlebige und benutzerfreundliche Standardsoftware für die Erstellung von Wörterbüchern aller Art. Sie ermöglicht es Ihnen, die Produktionszeit Ihrer Wörterbücher zu reduzieren und gleichzeitig deren Qualität und Konsistenz zu verbessern“ (<http://tshwanedje.com/deu/>, letzter Abruf 27.09.2021).

Da die Darstellungen und Beschreibungen von Lemmata in den heutigen Wörterbüchern immer vielfältiger und vernetzter werden, wachsen die Anforderungen an die Datenbanken exponentiell. In diesem Sinne wussten wir, dass einer der Vorzüge der TLex-Anwendung darin besteht, dass sie die Möglichkeit bietet, den DTD-Editor (Document Type Definition) individuell anzupassen. Es handelt sich um ein auf XML-Standards (eXtensible Markup Language) basierendem Tool, das an jedes Projekt angepasst werden kann und somit Lemmata auf flexible Weise darstellt und beschreibt. Die komplexe Informationsstruktur für jedes Lemma und die Anzeige seiner Bedeutungen, die wir in der lexikografischen Ressource anbieten wollten, führte zu Komplikationen bei der technischen Realisierung der Erstellung und Änderung der DTD. Die lexikologischen Daten sind alle auf inter- und intralingualer Ebene eng miteinander verknüpft. Diese Tatsache und der onomasiologische Ansatz erschwerten sowohl die Strukturierung der Informationen im Prozess als auch ihre Darstellung in der Ressource selbst (vgl. Abschnitt 4.3). Die Lösung dieses Problems war das Ergebnis intensiver gemeinsamer Arbeit zwischen den Hauptforschern des Projekts und dem beauftragten Unternehmen, das z.Z. der Erstellung dieses Beitrages für eine Reihe von Arbeiten beauftragt ist, welche sich noch „in progress“ befinden. Die zahlreichen Sitzungen waren nicht immer produktiv. Einerseits konnten wir unsere Vorstellungen und Bedürfnisse den Technologieexperten nicht immer eindeutig übermitteln und andererseits konnten wir nicht immer die technologischen Limitationen nachvollziehen¹⁵.

Zwischen der notwendigen manuellen Analyse der Daten, die in den Annotations-/Kodiertabellen erfasst wurden und der Übertragung dieser Daten in die TLex-Datenbank sollte eine Schnittstelle entwickelt werden, über

¹⁵ Siehe die nachdenklichen Beobachtungen von Tarp zu den Schwierigkeiten, die mit der Kommunikation zwischen Lexikografen und Programmierern bei der Entwicklung eines elektronischen Wörterbuchs verbunden sind, im Hinblick auf die *lexicographer's interface* und die Benutzerschnittstelle (Tarp, 2015, S. 230-241).



die diese Daten möglichst automatisch importiert werden können, um weitere aufwändige händische Arbeiten zu vermeiden. Dies konnte jedoch nur zum Teil realisiert werden (z.B. für die Belege) und nur nach einer weiteren komplexen manuellen Eingabe der vorstrukturierten Daten nach Lesarten in die DTD erfolgen.

Da natürlich auch die Gestaltung einer geeigneten Benutzeroberfläche mit den komplexen Zugriffsmöglichkeiten zu dem meist vielseitig vernetzten Informationsangebot direkt an die Information gekoppelt ist, die in der Datenbank im Hintergrund zur Verfügung steht, war ein ständiger Austausch zwischen den verantwortlichen Personen für die Datenanalyse und die sprachtechnologische Datenmodellierung notwendig. Die Projektleiterinnen sahen sich immer wieder in Situationen, in denen die Durchführung von bestimmten Vorstellungen nicht erwartbar hohe Kosten verursachten, die im Gesamtbudget des Projekts zu dem Moment der Planung und Antragstellung nicht genau veranschlagt werden konnten. So mussten immer wieder pragmatische Kompromisslösungen zwischen den Vorstellungen, wie eine multilinguale Ressource idealerweise sein sollte und den realen Möglichkeiten gesucht werden.

3.3. Herausforderungen bei der Erstellung der Benutzeroberfläche

Um den aktuellen Anforderungen an eine moderne Internetressource gerecht zu werden, stand die Entwicklung der Oberfläche der Ressource v.a. in Verbindung mit der Benutzeradaptivität und der Benutzerfreundlichkeit und umfasst u.a. folgende Aspekte:

(i) Sprachauswahl: Die anvisierte Zielgruppe der Ressource sind Lernende einer der drei Sprachen als Fremdsprache ab B2 Niveau, die Information in L2-Produktionssituationen brauchen. Unser Ziel war es, sowohl die konkrete Benutzungssituation als auch die Muttersprache der Benutzenden zu berücksichtigen. Es sollte hauptsächlich die Möglichkeit geschaffen werden, über unterschiedliche Sprachen auf die Ressource zuzugreifen. Der Zugriff über die einzelnen Sprachen steht allerdings nicht unbedingt in Verbindung mit der Muttersprache der Benutzenden, sondern mit der Fremdsprache, in der der Benutzende Ausdrucksmöglichkeiten sucht. Ein Zugriff über die Information in der Muttersprache ist zwar auch möglich, führt aber dann nur über das kontrastive Modul zu möglichen Entsprechungen in der Fremdsprache.

(ii) Allgemeine Metasprache: Die Herausforderung bestand darin, sich für eine Metasprache zu entscheiden, die für Benutzende mit unterschiedlichen Ausgangssprachen gleichermaßen verständlich ist, da nicht für alle drei Sprachen alle lexikografischen Angaben dreisprachig angeboten werden konnten. Dabei mussten nicht nur terminologische Diskrepanzen überwunden werden, die sich automatisch aus den unterschiedlichen grammatikalischen Ansätzen der einzelnen linguistischen Traditionen ergeben, sondern v.a. die eingeschränkte Sprachkompetenz der Benutzenden in den jeweiligen Fremdsprachen der Ressource. Unabhängig davon, ob es sich um Abkürzungen für syntaktische Bezeichnungen abstrakter Formeln ($S_P(+T)(B) = \text{SCHREIBER_PRODUKT}(+\text{THEMA})(\text{BENEFACTIV})$), für



semantische Rollen (ESC = SCHREIBER etc.) oder um Erklärungen, Definitionen oder Kommentare handelt, waren mehrere Arbeitsphasen notwendig, um uns letztendlich auf eine sprachlich adäquate Formulierung zu einigen. Bei der Wahl von Abkürzungen und abstrakten Formeln wurde für alle drei Sprachen eine identische Form gewählt, für die es Entschlüsselungshilfen über das dreisprachige Glossar und/oder über im Text selbst angebotene Erklärungen gibt.

(iii) Sprachliche Komplexität der Korpusbelege: Die Ressource bietet einen Zugriff auf alle Korpusbelege an, die die Basis der Analyse bilden. Da es sich um authentische Belege aus verschiedenen Textgenres handelt, ist das Verständnis dieser Belege bei L2-Lernenden ab dem B2-Niveau nicht immer gegeben. Eine Lösung dieses Problems stellt das Angebot von einigen Kompetenzbeispielen/Musterbeispielen für jede Lesart dar, welche über die Information zur KOMBINATORIK im linken Feld einsehbar sind (vgl. Abbildung 10 in Abschnitt 3.2. zur Kombinatorik von *schreiben 2*).

(iv) Metasprache vs. Objektsprache: Der konzeptuell-onomasiologische Zugriff erfordert die Versprachlichung der Konzepte. Die dafür verwendete Metasprache fällt häufig mit der zu beschreibenden Objektsprache zusammen. Mit welchen Mitteln können wir z.B. das Paradigma der KOMMUNIKATION von dem der PRODUKTION differenzieren? Bei der Reduzierung auf sprachliche Ausdrucksmittel, die eventuell einen hohen Abstraktionsgrad aufweisen, stoßen wir auf Verständigungsprobleme bei den anvisierten Benutzenden. Die Verwendung von nicht sprachlichen Mitteln, wie z.B. von visuellem Material war in unserem Projekt aus Kostengründen nicht möglich.

(v) Komplexe Informationsstruktur: Die Benutzenden sollten idealerweise über wenige Klicks zu der gesuchten Information gelangen. Die komplexe interne Struktur der Ressource CombiDigiLex (vgl. Abschnitt 3) erfordert jedoch eine Vielzahl von Klicks zur Auswahl der gewünschten Information. Zur Vermeidung von Rechercheabbrüchen war es daher erforderlich, eine komplexe Verlinkungsstruktur anzubieten, die es dem Benutzenden immer wieder ermöglicht, zum Ausgangspunkt zurückzugehen und/oder die Recherche zu korrigieren/modifizieren. Auf der Benutzeroberfläche ist dies durch Navigationshilfen gekennzeichnet.

(vi) Onomasiologische Zugriffsstruktur auf das Informationsangebot: Die Konzipierung einer onomasiologischen Zugriffsstruktur für Produktionszwecke im L2-Bereich erfordert nicht nur innovative Ideen bei der entsprechenden Ausführung, sondern auch ein Umdenken der Benutzenden. Im Allgemeinen sind Wörterbuchbenutzer vertraut mit semasiologisch-alphabetisch strukturierten Ressourcen, nicht aber mit Ressourcen, die von Konzepten und Ideen ausgehen. Idealerweise wäre ein konzeptuell-onomasiologischer Zugriff über visuelles Material eine geeignete intuitive Art für Produktionssituationen, auf Lexikalisierungsmöglichkeiten zuzugreifen. Wie

schon in (iv) erwähnt wurde, stand die Durchführung dieser Ideen nicht in einem akzeptablen Kosten-Leistungsverhältnis.

Zusammenfassung, Ausblick und Desiderata

Wie bekannt ist, ist die lexikografische Arbeit ein dynamischer und unendlicher Prozess, der eine permanente Aktualisierung, Reformulierung und Adaptation an innovative Technologien erfordert. Dies wird besonders deutlich, wenn man eine elektronische Ressource erstellt. Trotz dieser Tatsache haben wir die Schwierigkeiten nicht gescheut und die Entwicklung von CombiDigiLex in Angriff genommen.

Die Summe der hier aufgezählten unterschiedlichen Herausforderungen in Verbindung mit der Erstellung einer Ressource und die Lösungsmöglichkeiten stehen in direkter Verbindung mit den begrenzten finanziellen und menschlichen Ressourcen, die ein Forschungsprojekt normalerweise zur Verfügung hat. Die Ergebnisse stehen immer in Verbindung mit notwendigen Kompromissen, die ausgehandelt werden müssen. In diesem Sinne spiegelt der Prototyp CombiDigiLex zum aktuellen Zeitpunkt die Ergebnisse einer Projektarbeitsphase wider, die in weiteren Phasen ausgebaut, verbessert und sogar verändert werden können.

Für eine kommende Projektphase sind v.a. folgende Ziele anvisiert:

- Erhöhung der Anzahl von Lemmata mit lexikografischen Einträgen in den drei behandelten Sprachen
- Erweiterung der Suchfunktionalitäten: Erweiterte Suche über das Rolleninventar und die abstrakten Argumentstrukturmuster
- Ausbau von externen Verlinkungen zu anderen lexikografischen Ressourcen und Korpora
- Erweiterung der Datengrundlage auf gesprochene Daten der Interaktion
- Testung des Informationsangebotes in realen L2-Situationen

Das Projektvorhaben hat sich als sehr nützlich und interessant für fortgeschrittene Produktionssituationen im L2-Bereich gezeigt und auch die entwickelte korpusbasierte Methodik zur Datengewinnung und -analyse hat sich als angebracht und durchführbar erwiesen. Außerdem konnten die technischen und technologischen Hürden der Datenmodellierung überwunden sowie ein realistischer Kompromiss für die Entwicklung einer geeigneten Benutzeroberfläche gefunden werden. Dennoch war es bis jetzt nicht möglich, über einen lexikografischen Prototyp hinaus eine Ressource mit einer relevanten Anzahl von Einträgen zu erstellen, die in realen Situationen dem L2-Lernenden wirklich von Nutzen sein könnte. Gründe dafür liegen hauptsächlich daran, dass wir erstens in unserem Team nicht auf die Mitarbeit von Experten in Sprachtechnologie zählen konnten und zweitens an den arbeitsaufwändigen händischen Analysen der Korpusbelege und der anschließenden komplexen Datenstrukturierung und -aufbereitung für das lexikografische Informationsangebot. Aus diesem Grunde sehen wir es für den weiteren Ausbau der Ressource CombiDigiLex sowie für alle ähnlichen Projekte als unabdingbar an, redaktionell bearbeitete Information mit automatisch annotierten Korpusdaten zu bereichern und teilweise sogar zu ersetzen. Die Nutzung von



entsprechenden Annotationstools und die Entwicklung von Werkzeugen zur automatischen Erstellung von Verbprofilen auf der Basis einer umfangreichen Datenmenge aus Korpora zu verschiedenen Textsorten kann nicht nur das Informationsangebot in den Ressourcen, sondern auch Möglichkeiten für benutzeradaptive Recherchen erhöhen.

Literatur

- Boas, Hans (2009). 3. Semantic frames as interlingual representations for multilingual lexical databases. In Hans Boas (Hrg.), *Multilingual FrameNets in Computational Lexicography: Methods and Applications* (S. 59-100). De Gruyter. <https://doi.org/10.1515/9783110212976.1.59>
- Egido Vicente, María (im Dr.). Argumentstruktur kontrastiv: eine empirische Studie im Rahmen der Kommunikationsverben im Deutschen und Spanischen. Der Fall der Direktive *fragen* und *preguntar*. In Meike Meliss, Inmaculada Mas Álvarez, Paloma Sánchez Hernández und Vanessa González Ribao (Hrg.), *Argumentstrukturmuster im Sprachvergleich. Korpusbasierte Studien zu Verben ausgewählter Paradigmen*. De Gruyter.
- Fernández Méndez, Manuel (im Dr.). Die bifokale Perspektive der Verben aus dem LEIHEN-Paradigma: Konvergenzen und Divergenzen im Kombinationspotenzial ausgewählter Elemente. In Meike Meliss, Inmaculada Mas Álvarez, Paloma Sánchez Hernández und Vanessa González Ribao (Hrg.), *Argumentstrukturmuster im Sprachvergleich. Korpusbasierte Studien zu Verben ausgewählter Paradigmen*. De Gruyter.
- Fernández Méndez, Manuel; Franco Barros, Mario; Sueiro Orallo, Irene (2016). El uso de los diccionarios de/para ELE y DaF: análisis de una encuesta desde una perspectiva contrastiva. In Ferran Robles i Sabater, Daniel Reimann und Raúl Sánchez Prieto (Hrgs.), *Sprachdidaktik Spanisch - Deutsch. Forschungen an der Schnittstelle von Linguistik und Fremdsprachendidaktik* (S. 73-93). Narr. <http://hdl.handle.net/10400.13/2718>
- Fillmore, Charles J.; Johnson, Christopher R.; Petruck, Miriam R. L. (2003). Background to FrameNet. *International Journal of Lexicography*, 16(3), 235–250. <https://doi.org/10.1093/ijl/16.3.235>
- Franco Barros, Mario; Meliss, Meike (im Dr.). Die Verben *schreiben/escribir* im Vergleich: empirische Studie zu Argumentstrukturmustern, Bedeutungen und Frames. In Meike Meliss, Inmaculada Mas Álvarez, Paloma Sánchez Hernández und Vanessa González Ribao (Hrgs.), *Argumentstrukturmuster im Sprachvergleich. Korpusbasierte Studien zu Verben ausgewählter Paradigmen*. De Gruyter.
- Fuertes Olivera, Pedro A. (Hrg.) (2018). *The Routledge Handbook of Lexicography*. Routledge.
- González Ribao, Vanessa (2015). Sobre algunos conflictos en la 'pre'-lexicografía: la selección de corpus para la elaboración de un diccionario conceptual alemán-español. In María José Domínguez Vázquez, Xavier Gómez Guinovart und Carlos Valcárcel Riveiro (Eds.), *Lexicografía de las lenguas románicas: Aproximaciones a la lexicografía contemporánea y contrastiva. Volumen II* (S. 247-270). De Gruyter. <https://doi.org/10.1515/9783110310337>
- González Ribao, Vanessa (2021). *Mediale Kommunikationsverben: Das Zusammenspiel von Verb- und Musterbedeutung im Sprachvergleich Deutsch-Spanisch*. De Gruyter. <https://doi.org/10.1515/9783110718676>
- González Ribao, Vanessa; Meliss, Meike (2015). Theoretische und methodologische Grundlagen eines onomasiologisch-konzeptuell orientierten Produktionswörterbuches für den zweisprachigen Lernkontext: Deutsch-Spanisch. In José-Antonio Calañas Contente und Ferran Robles i Sabater (Hrgs.), *Die Wörterbücher des Deutschen: Entwicklungen und neue Perspektiven* (S. 109-136). Peter Lang. <https://www.peterlang.com/view/title/18070>



- González Ribao, Vanessa; Meliss, Meike (im Dr.). Argumentstrukturen in Theorie und Anwendung: Forschungsfragen im inter- und intralingualen Kontext. In Meike Meliss, Inmaculada Mas Álvarez, Paloma Sánchez Hernández und Vanessa González Ribao (Hrsg.), *Argumentstrukturmuster im Sprachvergleich. Korpusbasierte Studien zu Verben ausgewählter Paradigmen*. De Gruyter.
- González Ribao, Vanessa; Meliss, Meike; Proost, Kristel (im Dr.). Argumentstrukturen kontrastiv: Methodologische Grundlagen für korpusbasierte quantitative und qualitative Studien. In Meike Meliss, Inmaculada Mas Álvarez, Paloma Sánchez Hernández und Vanessa González Ribao (Hrsg.), *Argumentstrukturmuster im Sprachvergleich. Korpusbasierte Studien zu Verben ausgewählter Paradigmen*. De Gruyter.
- González-Ribao, Vanessa; Proost, Kristel (2015). El campo léxico al servicio de la lexicografía: Un análisis contrastivo en torno a algunos subcampos de los verbos de comunicación en alemán y español. In María José Domínguez Vázquez, Xavier Gómez Guinovart und Carlos Valcárcel Riveiro (Hrsg.), *Lexicografía de las lenguas románicas: Aproximaciones a la lexicografía contemporánea y contrastiva. Volumen II* (S. 223-245). De Gruyter. <https://doi.org/10.1515/9783110310337>
- Gouws, Rufus H.; Heid, Ulrich; Schweickard, Wolfgang; Wiegand, Herbert E. (2013). *Dictionaries. An International Encyclopedia of Lexicography: Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. De Gruyter. <https://doi.org/10.1515/9783110238136>
- Granger, Sylviane; Paquot, Magali (Hrsg.) (2012). *Electronic Lexicography*. Oxford University Press.
- Klosa, Annette; Müller-Spitzer, Carolin (2016). *Internetlexikografie: Ein Kompendium*. De Gruyter. <https://doi.org/10.1515/9783050095615>
- Mas Álvarez, Inmaculada; Santalla del Río, María Paula; Iglesias Cancela, Yolanda (2017). La enseñanza de la lexicografía del español en la universidad: una renovación necesaria. *Cauce: Revista Internacional de Filología, Comunicación y sus Didácticas*, 40, 17–27. <https://hdl.handle.net/11441/81991>
- Mas Álvarez, Inmaculada; Santos Palmou, Xandra (im Dr.). Das kombinatorische Profil der Verben (*sich*)*fühlen*, *sentir(-se)*_{PT} und *sentir(se)*_{SP}. Ein multilingualer Vergleich von ihrer Verwendung in journalistischen Texten. In Meike Meliss, Inmaculada Mas Álvarez, Paloma Sánchez Hernández und Vanessa González Ribao (Hrsg.), *Argumentstrukturmuster im Sprachvergleich. Korpusbasierte Studien zu Verben ausgewählter Paradigmen*. De Gruyter.
- Meliss, Meike (2014). Die fremdsprachige Produktionssituation im Fokus eines onomasiologisch-konzeptuell orientierten, zweisprachig-bilateralen Wörterbuches für das Sprachenpaar Deutsch–Spanisch: Theoretische und methodologische Grundlagen von DICONALE. In Andrea Abel, Chiara Vettori und Natascia Ralli (Hrsg.), *Proceedings of the XVI EURALEX International Congress: The User in Focus* (S. 1119-1134). Institute for Specialised Communication and Multilingualism. https://www.euralex.org/elx_proceedings/Euralex2014/euralex_2014_087_p_1119.pdf
- Meliss, Meike (2015). Was suchen und finden Lerner des Deutschen als Fremdsprache in aktuellen Wörterbüchern? Auswertung einer Umfrage und Anforderungen an eine aktuelle DaF-Lernerlexikographie. In Thorsten Roelcke (Hrsg.), *Wörterbücher für Deutsch als Fremdsprache –Probleme und Perspektiven* (S. 401-432). Info Daf Themenheft 4/2015. <http://hdl.handle.net/10347/20328>
- Meliss, Meike; Egido Vicente, María; Fernández Méndez, Manuel (2019). Plädoyer für die Entwicklung einer digital-lexikografischen Kompetenz im Fremdsprachenunterricht. *Lexicographica*, 34(2018), 123-156. <https://doi.org/10.1515/lex-2018-340107>
- Meliss, Meike; Mas Álvarez, Inmaculada; Sánchez Hernández, Paloma; González Ribao, Vanessa (Hrsg.), *Argumentstrukturmuster im Sprachvergleich. Korpusbasierte Studien zu Verben ausgewählter Paradigmen*. De Gruyter.



- Meliss, Meike; Sánchez Hernández, Paloma (im Dr.). Das Konzept LANGEWEILE im Deutschen und Spanischen: eine korpusbasierte inter- und intralinguale Studie zu verbalen Argumentstrukturmustern. In Meike Meliss, Inmaculada Mas Álvarez, Paloma Sánchez Hernández und Vanessa González Ribao (Hrsg.), *Argumentstrukturmuster im Sprachvergleich. Korpusbasierte Studien zu Verben ausgewählter Paradigmen*. De Gruyter.
- Nied Curcio, Martina (2013). Der Gebrauch zweisprachiger Wörterbücher aus der Sicht italienischer Germanistikstudierender. *Lexicographica*, 29, 129-145. <https://doi.org/10.1515/lexi-2013-0009>
- Runte, Maren (2015). *Lernerlexikographie und Wortschatzerwerb*. De Gruyter. <https://doi.org/10.1515/9783110428476>
- Taborek, Janusz (2019). Wörterbuchbenutzung von polnischen Germanistikstudierenden. *Lexicographica*, 34(2018), 207-222. <https://doi.org/10.1515/lex-2018-340110>
- Tarp, Sven (2015). Structures in the communication between lexicographer and programmer: Database and interface. *Lexicographica*, 31(1), 217-246. <https://doi.org/10.1515/lexi-2015-0011>
- Trawiński, Beata; Kupietz, Marc (2021). Von monolingualen Korpora über Parallel- und Vergleichskorpora zum Europäischen Referenzkorpus EuReCo. In Henning Lobin, Andreas Witt und Angelika Wöllstein (Hrsg.), *Deutsch in Europa. Sprachpolitisch, grammatisch, methodisch. Jahrbuch des Instituts für Deutsche Sprache 2020* (S. 209-234). De Gruyter. <https://doi.org/10.1515/9783110731514-012>

