

Zita Hollós

Cross-Media-Publishing in der korpusgestützten Lernerlexikographie

Entstehung eines Lernerwörterbuchportals DaF

Abstract This paper gives an insight into a cross-media publishing process on different stages: from a printed bilingual syntagmatic dictionary for GFL to an online learner's dictionary of German collocations to a German learner's dictionary portal. On the basis of an sql database specially developed for a corpus-guided dictionary of German collocations, the bilingual syntagmatic learner's dictionary KOLLEX was published in 2014. The first part of the article describes this lexicographic process, focusing the most relevant aspects of the dictionary concept, e.g. dictionary type, subject matter, corpus-guided data selection and microstructure. The second part introduces the first online version of KOLLEX from 2016 and the profound changes in the editing system – from a desktop version (2005) to a web-based editing system (2016) –, which resulted successively in a prototype of a German learner's dictionary portal, called E-KOLLEX DAF (2018-). Focusing on the aspects of dynamism and integration of different resources from a learner's perspective the paper shows the innovative features of this new online reference work. The contribution presents the solutions for the integration of new datatypes in the database of KOLLEX and the linking to different data in German monolingual dictionary platforms. The paper outlines the web design, functioning and technical improvements of E-KOLLEX DAF. The conclusions provide an outlook to the forthcoming challenges.

Keywords Wörterbuchportal, Kollokationen, Lernerwörterbuch, DaF, Datenbank, Webdesign

1. Etappe I: KOLLEX DAF, das Printwörterbuch (2004-2014)

1.1 Datenbankbasiertes und korpusgestütztes Wörterbuchprojekt für Kollokationen

Der erste Band des Nachschlagewerks „Wörterbuch zur Lexikographie und Wörterbuchforschung“ (WLWF) definiert *Cross-Media-Publishing* folgendermaßen: „Publikationsprozess, der darauf hin konzipiert ist, aus ein- und derselben Datenbasis Produkte für beliebige Publikationsmedien zu erzeugen“ (WLWF 2010, S. 751). Im vorliegenden Beitrag wird dementsprechend das Ziel verfolgt, die Ergebnisse eines solchen Prozesses zu präsentieren: wie auf der Grundlage einer Datenbank für deutsch-ungarische Kollokationen ein syntagmatisches Printwörterbuch DaF publiziert wurde und daraus wenig später ein Internetwörterbuch für Kollokationen bereits für ein breites Lernerpublikum entstand. Gleichzeitig wird der Artikel einen ersten Überblick über die Etappen eines datenbankbasierten lexikographischen Prozesses mit geplanter Print- und dann digitaler Publikation geben. Wie letztendlich ein Prototyp eines Wörterbuchportals DaF sukzessive auf dieser Datengrundlage entwickelt wurde, konnte bisher umfassend nicht erörtert werden. Der vorliegende Beitrag will auch diese Lücke schließen.

1.2 Zweisprachiges syntagmatisches Printwörterbuch DaF

Moderne Korpus- und Sprachtechnologie sowie Verwaltung und Modellierung von Daten in einer Datenbank sind heutzutage beim Aufbau verschiedener Sprachressourcen ebenso wenig wegzudenken wie die Datenpräsentation, also die Planung einer dynamischen Benutzeroberfläche und eines attraktiven Webdesigns (vgl. Klosa/Müller-Spitzer 2016).

Als 2005 das Forschungs- und Wörterbuchprojekt SZÓKAPTÁR/KOLLEX (im Weiteren KOLLEX) mit dem Ziel der Erstellung eines korpus- und datenbankbasierten zweisprachigen Kollokationswörterbuchs für Deutschlerner gestartet wurde, stand die Korpus- und Sprachtechnologie für das Deutsche noch relativ am Anfang. Die Realisierung datenbankbasierter elektronischer Wörterbücher – die im Rahmen einsprachiger Ressourcen wie *ellexiko* bereits erprobt wurde (vgl. Hass 2005) –, war ein komplett neues Terrain für die zweisprachige Speziallexikographie, insbesondere für die Kollokationslexikographie für Sprachlerner.

In diesem Kapitel werden die wichtigsten Phasen eines speziallexikographischen Projekts vorgestellt, dessen Ergebnisse sowohl als Printpublikation (Hollós 2014) als auch teilweise in digitaler Form bereits im Jahre 2016 vorlagen. Seitdem entsteht sukzessive ein Wörterbuchportal für ungarischsprachige Deutschlerner.¹

1.2.1 Forschungsprojekt für ein syntagmatisches Lernerwörterbuch

Die theoretischen Vorarbeiten lagen bereits in Form einer Lernerwörterbuchkonzeption (vgl. Hollós 2004) vor. Ohne diese Planungsphase (2000-2003) wäre die Phase der Datenmodellierung und -speicherung in einer sql-Datenbank sowie die Erstellung eines umfangreichen, zweisprachigen, korpusgestützten Lernerwörterbuchs der Kollokationen für Deutschlerner (2004-2014) nicht möglich gewesen.

Für das durch den *Ungarischen Nationalfonds für Wissenschaft und Forschung* (OTKA) von 2005-2008 finanzierte Forschungsprojekt wurde eigens ein Datenbank- und Redaktionssystem mit vielen modernen Funktionen für das geplante Spezialwörterbuch entwickelt. Im Rahmen der lexikographischen Werkstatt der Universität wurden ein Informatiker, im Durchschnitt drei studentische Mitarbeiter und zwei deutsche Muttersprachler als Mitarbeiter und Lektoren beauftragt. Die Korpusdaten für das Kollokationswörterbuch sind der Zusammenarbeit mit dem Leipziger Projekt *Deutscher Wortschatz* (DW) zu verdanken.

Als Endprodukt ist ein zweisprachiges, syntagmatisches Lernerwörterbuch für Deutschlerner der Mittel- und Oberstufe entstanden, das insgesamt mehr als 61 000 Kollokationen, Wortverbindungen, Kombinationen mit Valenzangaben enthält, von denen mehr als 48 000 auch ins Ungarische übersetzt wurden. Im KOLLEX sind somit mehr als 10 000 verschiedene Lexeme als Kollokatoren/Kotextpartner zu 2262 Basis-Lemmata verzeichnet. Die exakten Zahlen des zweisprachigen, syntagmatischen Printwörterbuchs SZÓKAPTÁR/KOLLEX präsentiert die Tabelle 1:

¹ E-KOLLEX (DAF) ist seit 2016 unter der folgenden Internetseite zu finden: www.kollex.hu/szotar.

Basislemmata (im Wörterverzeichnis)	2262
Kollokatoren/Kotextpartner (im Register)	10313
Kollokationen/Wortverbindungen (nach Strukturtypen)	48757
Kombinationen (eigenständiger Strukturtyp für komplexe Kollokationen)	2661
Valenzrealisierungen (bei Verblemmata)	8590
Basen zu den Kollokatoren (bei Adjektivlemmata) ²	1609
Kollokationen/Kotextpartner insgesamt	61617

Zahlen von SZÓKAPTÁR/KOLLEX

Tabelle 1: Exakte Zahlen von Lemmata, Kollokationen/Wortverbindungen, Kombinationen u.a.

1.2.2 Wörterbuchtyp, Wörterbuchgegenstand und Datenselektion

Im Folgenden werden ausgewählte Aspekte der Wörterbuchkonzeption, wie der Wörterbuchtyp, der Wörterbuchgegenstand sowie die korpusgestützte Datenselektion (i.S.v. Klosa 2011, S. 112) erörtert.³

Die ursprüngliche Wörterbuchcharakteristik (vgl. Hollós 2004, S. 174) musste im Hinblick auf das Ergebnis des publizierten Printwörterbuchs hin modifiziert werden: KOLLEX (2014) ist ein korpusgestütztes und datenbankbasiertes, polyfunktionales, dennoch produktionsbezogenes zweisprachiges syntagmatisches Lernerwörterbuch, dessen primärer Wörterbuchgegenstand die deutschen Kollokationen zu den Autosemantika der Wortliste des Zertifikats Deutsch⁴ umfasst.

Im KOLLEX wird mit einem neuentwickelten Kollokationsbegriff gearbeitet. Die fremdsprachendidaktisch motivierte Kollokationsauffassung (i.S.v. Hausmann 1984) wurde im Rahmen eines integrativen Kollokationsbegriffs mit den Ergebnissen der Sprachtechnologie (vgl. Quasthoff 2009) kombiniert.⁵ Statistische Kookkurrenzprofile wie auch im DWDS-Wortprofil (vgl. Geyken 2011) liefern große Mengen an sprachlichen Daten für statistisch signifikante Kookkurrenzpartner. Viele sind i.o.S.

² Mit der Umkehrung der gerichteten Kollokationsrelation bei den Lemmzeichentypen Adjektiv und Verb wurden neue Datentypen in den Wörterbuchartikeln etabliert: Bei den adjektivischen Lemmata – in diesem Fall nicht als Basis, sondern als Kollokator verstanden – werden auch Kollokationsbasen, also typische Substantive und Verben aufgelistet. Ebenfalls kann man die substantivischen Valenzrealisierungen als Basen zum Verb (Kollokator) interpretieren.

³ Aus Platzgründen beschränke ich mich auf eine Zusammenfassung der Ergebnisse des Wörterbuchprojekts zwischen 2004-2014 und verweise auf die Projekthomepage mit einzelnen Publikationen zu spezifischen Aspekten der Wörterbuchkonzeption und -arbeit: www.kollex.hu.

⁴ Einzelheiten zur Lemmaselektion sind in Hollós (2004, 154-159) zu finden.

⁵ Näheres zu diesem Kollokationsbegriff findet man in Hollós (2016). Die Unterscheidung in intra- und interlinguale Kollokationen integriert auch den kontrastiven Aspekt und bildet damit die theoretische Grundlage für das ursprünglich geplante zweite Wörterverzeichnis der intralingualen Kollokationen, deren automatische Generierung die alte Desktop-Version des Programms noch immer ermöglicht. Das Printwörterbuch enthält außerdem 8378 mit dem Symbol „^d“ markierte, interlinguale Kollokationen.

zur Basis affine Kollokatoren. Ausgehend von der Basis, deren Sememstruktur für alle Wörterbucheinträge als erstes erarbeitet werden musste, können mit (beinahe) muttersprachlicher Kompetenz typische Kollokatoren/Kotextpartner sememspezifisch ausgewählt und übersetzt, bzw. mit stilistisch-pragmatischen Angaben sowie mit morphosyntaktischen Restriktionen versehen werden.

Da zur Zeit der Datenselektion und -aufbereitung das DWDS-Wortprofil noch nicht zur Verfügung stand, wurden die Korpusdaten für das Wörterbuch in Zusammenarbeit mit dem Leipziger Korpus- und Wortschatzprojekt gewonnen.⁶ Darüber hinaus wurde die am IDS entwickelte *Kookkurrenzdatenbank* (CCDB) von Belica (1995) genutzt sowie Daten aus den damals modernsten ein- und zweisprachigen Wörterbüchern übernommen. Das syntagmatische Lernerwörterbuch KOLLEX ist dementsprechend korpusgestützt (i.S.v. Klosa 2011, S. 112)⁷ erarbeitet worden, da bei der Selektion und Anordnung der Kollokatoren/Kollokationen korpusgesteuert, bei der Angabe der Valenz und der Auswahl der Valenzrealisierungen bei den Verblemmata korpusbasiert, mit den Daten von VALBU gearbeitet wurde.⁸

1.3 Artikelstruktur im KOLLEX

Die folgende Abbildung zeigt einen typischen Wörterbuchartikel aus dem KOLLEX:⁹

E-Mail [ˈi:meɪl] die / dn/oszt/sv das, der E-Mail/des
E-Mails, die E-Mails <fn> **e-mail**
SUBS **eine Menge E-Mails** † rengeteg/egy csomó
e-mail
ADJ **dringend** sürgős • csak jel **unerwünscht**
kéretlen, nem kívánatos • **formlos** nem hi-
vatalos, kötetlen formájú • **unpersönlich**
személytelen • **infiziert** vírusos
VERB **kissé vál** **senden**^{AKK} (el)küld • **bekommen-**
^{AKK} kap • **speichern**^{AKK} (el)ment • **lö-**
schen^{AKK} töröl • **weiterleiten**^{AKK} továbbít
• **vál archivieren**^{AKK} megőriz, archivál •
markieren^{AKK} pl. továbbításra, törlésre kijelöl
• **auf eine E. antworten** † válaszol egy
e-mailre • **etw**^{AKK} **per E. bestellen** † e-ma-
ilben rendel (meg) vmit • **etw**^{AKK} **per E. ver-**
breiten † e-mailben terjeszt vmit
KOMB **eine dringend zu beantwortende E.** sür-
gösen megválaszolandó e-mail

⁶ Die Analyse erfolgte mit dem Log-Likelihood-Maß von Dunning (vgl. Quasthoff 2009, S. 159).

⁷ Annette Klosa (2011, S. 110ff.) verwendet den Terminus „korpusgestützt“ oberbegrifflich für „korpusvalidierend“ (engl. corpus-based) und „korpusgesteuert“ (engl. corpus-driven).

⁸ Näheres findet man zur korpusgesteuerten und -basierten Datenselektion und -aufbereitung in Bezug auf die Verblemmata in Hollós (2008).

⁹ Die online Version des Wörterbuchartikels findet man in Abbildung 2.

wa₁ zu **E-Mail** im KOLLEX

Abb. 1: Darstellung der Artikelstruktur zum Lemmazeichen E-Mail, Abkürzungen: dn/oszt/sv=sdt./österr./schweiz., fn=Substantiv, csak jel=nur attributiv, kissé vál=ein wenig gewählt

Im Umtext „Kurze Benutzungshinweise“ findet man eine detaillierte Beschreibung der Artikelstruktur, deren Anfang hier wiedergegeben wird:

Der zentrale Gegenstand des Wörterbuchs sind Kollokationen und typische Wortverbindungen zu den einzelnen Stichwörtern. Sie werden in den Wörterbuchartikeln wortartspezifisch, nach den Strukturtypen der Kollokationen angeordnet:


zu den Substantiven **SUBSTANTIVE**, **ADJEKTIVE** und **VERBEN**

zu den Verben **ADVERBIEN**

zu den Adjektiven (und Adverbien) **ADVERBIEN**.

Teils werden nur die obigen Partner/Kollokatoren (bei Verben oft mit der jeweiligen Valenz) genannt, teils auch das abgekürzte Stichwort mitverzeichnet und damit die ganze Wortverbindung/Kollokation angegeben. [...]

Bei allen Lemmazeichentypen kann die Kategorie der **KOMBINATIONEN** auftreten. Hier werden größere, schwach oder teil-idiomatische Wortverbindungen verzeichnet. Das abgekürzte Lemma ist bei diesen Kombinationen obligatorischer Teil der Verbindung. (KOLLEX 2014, S. 963)

Außerdem ist hervorzuheben, dass die Kollokationen/Wortverbindungen mit ihren morphosyntaktischen Restriktionen (z. B.: csak jel=nur attr) und auch mit stilistisch-pragmatischen Angaben (z. B.: kissé vál=ein wenig gewählt) versehen sind. Didaktisch relevant ist des Weiteren, dass bei verbalen Kollokatoren auch die sog. Partnerfunktion (z. B.: AKK) als Index angegeben wird und die interlingualen Kollokationen mit dem Symbol „“ markiert sind (siehe Abb. 1).¹⁰

2. Etappe II: E-KOLLEX DAF, das Lernerwörterbuchportal (2015-2020)

2.1 Internetwörterbuch für Kollokationen

Im ersten Band des WLWF (2010) findet man zum Terminus *Cross-Media-Publishing* weitergehende Erklärungen, die den Begriff näher spezifizieren:

Im Anwendungsbereich der Lexikographie meint *Cross-Media-Publishing*, dass aus derselben Datenbasis verschiedene lexikographische Produkte (Printwörterbuch, CD-ROM-Wörterbuch, Online-Wörterbuch) erzeugt werden können. Während der Terminus *Cross-Media-Publishing* insbesondere auf die Spezifika einzelner Publikationsmedien abhebt, perspektiviert der bedeutungsverwandte Terminus *medienneutrales Publizieren* die Datenaufbereitung und -haltung, bei der das Datenmodell möglichst unabhängig von der Präsentation der Daten in verschiedenen Publikationsmedien bleiben sollte. (WLWF 2010, S. 751)

¹⁰ Alle Charakteristika des Printwörterbuchs sind im deutschsprachigen Nachspann (KOLLEX 2014, S. 953-979) ausführlich beschrieben.

Zum Online-Release der Daten wurde das Redaktionssystem im Jahre 2016 von einer Desktop-Applikation auf ein web-basiertes User-Interface umgestellt und gleichzeitig in Teilen als erster Prototyp eines zweisprachigen Internetwörterbuchs für Kollokationen für das breite Lernerpublikum kostenlos und ohne Werbung ins Netz gestellt. Die automatische Verlinkung mit bereits existierenden deutschen wissenschaftlichen Sprachressourcen hat mehrere Vorteile für ein online Kollokationswörterbuch: größere Zuverlässigkeit der Daten, Ökonomie im Hinblick auf sprachliche Ressourcen, gesteigerte Interaktivität durch durchdachte Vernetzung, gezielte komplementäre Effekte sowie Synergieeffekte zwischen den Ressourcen.

Dank der kontinuierlichen Weiterentwicklung seit 2016 ist allmählich ein Wörterbuchportal (i.S.v. Engelberg/Storrer 2016, S. 57) für Lerner entstanden, das im nächsten Abschnitt vorgestellt wird.

2.2 Deutsches Lernerwörterbuchportal

Der erste Prototyp wurde vor dem Hintergrund der Ergebnisse der Wörterbuchbenutzungsforschung für Onlinewörterbücher (vgl. Müller-Spitzer/Koplening/Töpel 2012) entworfen.¹¹ Im Folgenden werden die bereits (teil)realisierten Module des E-KOLLEX DAF in Kürze vorgestellt.

Die ersten zwei Module *KOLLEX* und *Schulwörterbuch* bilden das lexikographische Grundgerüst des Lernerwörterbuchportals, weil sie alle 1274 substantivischen Wörterbuchartikel aus dem Printwörterbuch SZÓKAPTÁR/KOLLEX (2014) und 693 printlayout-formatierte Substantivartikel des Schulwörterbuchs HOLL-SULI (2001) umfassen (siehe Tab. 2). Das zweisprachige Grundschulwörterbuch HOLL-SULI wurde sukzessive retrodigitalisiert, damit lernerorientierte Beispielsätze und ihre jeweiligen Übersetzungen zu den – auch im KOLLEX vorhandenen – Lemmata online dargeboten werden können. Da Kollokationen in der Sprachproduktion eine zentrale Rolle spielen (vgl. Réder 2006) und in einem Lernerwörterbuchportal zum primären Wörterbuchgegenstand gehören sollten, bilden die Daten von KOLLEX das Herzstück des Portals. Die anhand des DW ermittelten, statistisch signifikanten und manuell sememspezifisch ausgewählten Kollokationen und lernerrelevanten Wortverbindungen mit ihren jeweiligen Übersetzungen stammen also aus der Datenbank von KOLLEX, wo sie nach ihren Strukturtypen geordnet und mit vielen anderen semantischen und pragmatischen Angaben gespeichert vorliegen. Sie werden in der Webseitenarchitektur unter den linear angeordneten Strukturlinks dargeboten, nebst dem ersten Tab namens *Suliszótár* (dt. *Schulwörterbuch*).

Das dritte Modul *externe Ressourcen* impliziert die Integration lemmabezogener Portale: Bei E-KOLLEX DAF sind diese das lexikologische Portal DW und das Wortschatzportal DWDS (vgl. Engelberg et al. 2016). Damit wird gleichzeitig teilweise das vierte Modul *Korpus mit spezifischen Analysetools* realisiert.

¹¹ Dieser Prototyp wurde in Hollós (2019) detailliert beschrieben.

Gouws (2014) stellt mit Hilfe von Screenshots von *elexiko* plastisch vor, wie dank der elektronischen Datenpräsentation eine neue Datendistribution und eine neue Artikelstruktur des Typs „dynamic multi-layered article structure“ (Gouws 2014, S. 165) entsteht:¹²

Clicking on the structural indicator „weiter“ (= further) in anyone of these sub-blocks guides the user to a further layer in the treatment of the lemma [...]. Clicking on the data-identifying entry “Grammar” guides a user to the display in Figure 5. The grammar data here would typically be part of a traditional comment on form but here it is isolated from the text block containing comment on form items [...]. (Gouws 2014, S. 164)

Diese Art von Links werden bei Engelberg et al. (2016, S. 159) als Struktur- und Inhaltslinks definiert. Erstere sind Elemente der inneren Zugriffsstruktur, die letzteren gehören jedoch zum Wörterbuchgegenstand.

Die neue Dynamik entsteht demgemäß auch im Portal E-KOLLEX DAF einerseits durch Strukturlinks und andererseits dank neuer Datentypen, realisiert durch Inhaltslinks mit verschiedenen typisierten Linkanzeigen, die auch den direkten Zugriff auf weitere Ressourcen ermöglichen.

Erstrangiges Ziel der folgenden Unterkapitel ist es dementsprechend, einige lexikographische Potenziale moderner Onlineressourcen (vgl. Lemberg 2001; Gouws 2014, Müller-Spitzer et al. 2014 u.a.), insbesondere die Dynamik am aktuellen Prototyp des E-KOLLEX-Portals mit seinen Modulen nachzuweisen.¹³

Diese Potenziale sind bei Lemberg (2001, S. 73) bereits früh aufgelistet und erörtert worden. Von ihnen werden hier nur drei genannt und im Folgenden exemplifiziert:

- Hypertextualisierung,
- multimediale Aufbereitung lexikographischer Daten,
- Aufhebung eines statischen zugunsten eines dynamischen Wörterbuchs.

2.2.1 Dynamik durch neue Datentypen

Die Umstellung des Datenbankmanagement- und Redaktionssystems von einer Desktop-Applikation (2005-2015) zu einem webbasierten, dynamischen System (2016-) hat das erste Onlinerelease der Daten von E-KOLLEX ermöglicht und alle neuen Anforderungen von Lemberg (2001) bis auf die Kooperation zwischen Lexikographen und Nutzern gleichzeitig erfüllt. Neue Datentypen, wie z. B. Bild- und Tondateien, konnten in die Datenbank integriert werden, die zur Multimodalität der Ressource beigetragen haben.¹⁴ Die neue Dynamik entstand auch durch Hypertextualisierung, indem Struktur- und Inhaltslinks im Webdesign realisiert wurden. Der nächste Screenshot zeigt diese Elemente auf der Präsentationsebene

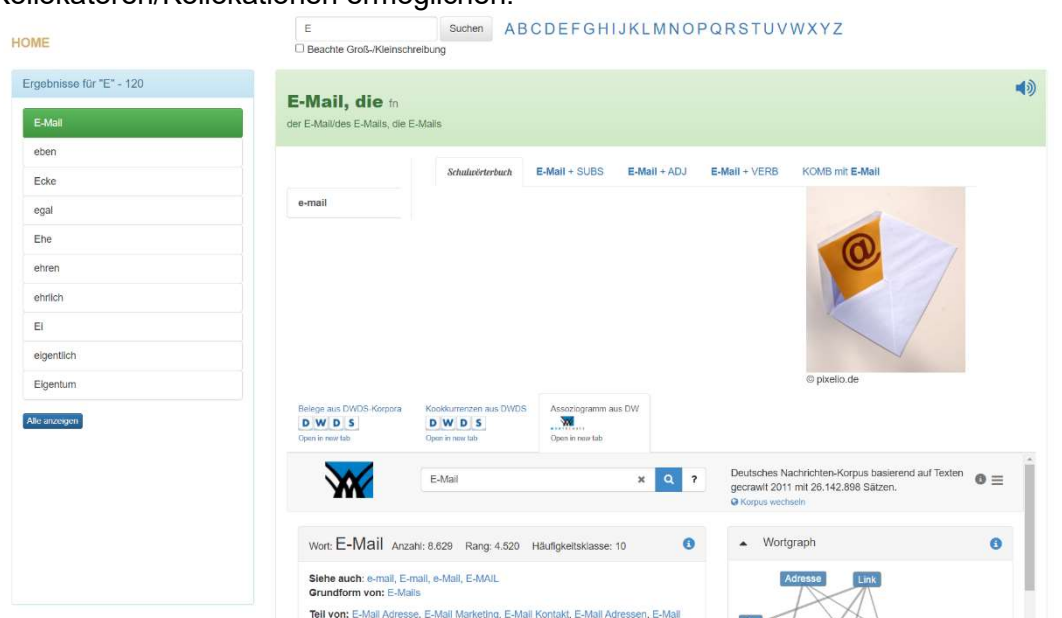
¹² Näheres zu dieser Datendistribution und Artikelstruktur im E-KOLLEX relativ zum Webdesign findet sich in Hollós (2018, 164-167).

¹³ Da hier Kürze geboten ist, siehe ausführlich Hollós (2018, 156-169).

¹⁴ Die Bereicherung des Internetwörterbuchs durch semenspezifische Bilder für die substantivischen Lemmata erfolgte systematisch am Anfang jeder Lemmastrecke. Dadurch gibt es fast 400 Bilder im Portal. Die Tondateien sind allerdings noch nicht realisiert.

Document template EURALEX

von E-KOLLEX DAF, indem rechts ein Bild zum Lemmazeichen und etwas rechts oberhalb ein Symbol für die Aussprache sowie oben und unten verschiedene Reiter zu sehen sind. Die unteren Tabs sind Inhaltslinks wie „Assoziogramm“ aus DW sowie „Belege“ und „Kookkurrenzen“ aus DWDS, während die oberen Strukturlinks gemäß der Strukturtypen der Kollokationen den gezielten Zugriff auf verschiedene Kollokatoren/Kollokationen ermöglichen:



Screenshot des Portals E-KOLLEX DAF mit dem Lemma **E-Mail**

Abb. 2: Neue Datentypen (Bild-, Tondatei) und verschiedene Tabs (unten: Inhaltslinks, oben: Strukturlinks) im Wörterbuchartikel zum Lemmazeichen E-Mail

Bei polysemen substantivischen Lemmazeichen wird in der ersten Lemmateilstrecke (hier: von **E-Mail** bis **Eigentum**) pro Semem jeweils ein Bild angesetzt (siehe z. B. Semem 2 zu **Nachricht** mit der Bedeutung 'kurze mündlich oder schriftliche Mitteilung' in Abb. 2). Das erleichtert einerseits die Desambiguierung des Wortes bei den Lernern und sorgt zugleich für ein noch dynamischeres multimodales Webdesign der Wörterbuchartikel.

2.2.2 Integration weiterer Ressourcen

Neue Perspektiven für das damalige Internetwörterbuch hat jedoch erst die Integration eines Schulwörterbuchs und die der bereits erwähnten einsprachigen lemmabezogenen Portale wie DW und DWDS eröffnet. Letztere sorgen für eine monodirektionale Vernetzungsstruktur mit (teilweise) direktem Zugriff auf bestimmte Datentypen in den ausgewählten Onlineressourcen (vgl. Hollós 2018). Gleichzeitig werden die Lemmazeichen mit Korpora verlinkt. Dies ermöglicht wiederum die gezielte Recherche nach bestimmten morphologischen, syntaktischen und semantischen Eigenschaften der nachgeschlagenen Wörter in Belegen.

Für die Integration des Schulwörterbuchs HOLL-SULI (2001), das für ungarischsprachige Deutschlerner zuerst als Printwörterbuch erschien, mussten alle Wörterbuchartikel zuerst aktualisiert werden, damit 2017 mit der Eingabe der Substantivartikel in die Datenbank begonnen werden konnte.

Der folgende Screenshot zeigt nicht nur den retrodigitalisierten Schulwörterbuchartikel, sondern unter Semem 2 auch einen weiteren Link „**Nachrichten**“, der für diesen Artikel aus dem PONS Bilderwörterbuch manuell ausgesucht und eingefügt wurde. Dieser Schulwörterbuchartikel bedeutet gleichzeitig den Einstieg in den Portalartikel zum Lemma **Nachricht**:

The screenshot shows a digital dictionary interface for the word 'Nachricht' (Hungarian: híre). On the left, there are three numbered sections: 1 - vkinek/vminek a híre, hire; 2 - üzenet; 3 - tévében v rádióban hírek. The main content area features a title 'die Nachricht {die Nachrichten}' and several numbered entries with examples in Hungarian. A small image of a smartphone with a message on the screen is visible on the right. At the bottom, there are two search bars: 'szinonimák(OpenThesaurus) érdekeségek(EyePloer)' and 'DaF-szócikkek(TheFreeDictionary) szócikk(elekiko)'.

wa₂ zu **Nachricht** im E-KOLLEX DAF

Abb. 3: Inhaltslinks zu weiteren Ressourcen im Wörterbuchartikel zum Lemmazeichen Nachricht

In der obigen Abbildung sind nicht nur die lernerrelevanten mikrostrukturellen Charakteristika des Schulwörterbuchs ersichtlich, sondern im Postkommentar auch weitere Inhaltslinks, deren Linkanzeiger auch typographisch hervorgehoben sind und in einer eigenen Suchzone erscheinen.¹⁵

Das zweisprachige Internetwörterbuch (2016) und das nachfolgende, seit 2017 im Entstehen befindliche Lernerwörterbuchportal E-KOLLEX DAF wurden von Anfang an für die deutschlernende oder sprachinteressierte Öffentlichkeit kostenlos unter der Internetadresse www.kollex.hu/szotar in großen Teilen zugänglich gemacht.

Die entsprechenden Kennzahlen des Lernerwörterbuchportals E-KOLLEX DAF findet man in der nächsten Tabelle (Stand März 2022):

Substantivlemmata	1274
Kollokationen/Wortverbindungen	ca. 30000
Wörterbuchartikel aus HOLL-SULI	693

¹⁵ Diese Inhaltslinks – ähnlich wie die Bilder – werden z. Z. nur bei den Musterartikeln, d.h. bei substantivischen Lemmazeichen in der ersten Lemmateilstrecke (hier: von **Nachbar** bis **Nagel**) angesetzt.

Bilder zu Sememen der Substantivlemmata	269
Dynamische Links zu Angaben in externen Ressourcen (DW, DWDS)	3822
Links zu weiteren externen Ressourcen (PONS-Bildwörterbuch etc.)	ca. 2700
Inhaltslinks insgesamt (DW, DWDS, PONS etc.)	ca. 6500

Zahlen von E-KOLLEX DAF

Tabelle 2: Exakte und geschätzte Zahlen von Substantivlemmata, Wortverbindungen/Kollokationen, Links, retrodigitalisierten Schulwörterbuchartikeln, Bildern u.a. im E-KOLLEX DAF

3. Herausforderungen der Etappe III von E-KOLLEX DAF

Anstatt eines Fazits wird in diesem letzten Kapitel der Versuch unternommen, einen Ausblick zu bieten, indem die möglichen Wege einer Weiterentwicklung des Portals E-KOLLEX DAF und die damit verbundenen Herausforderungen skizziert werden. Manche Desiderata – wie die Erweiterung der Datenbank mit Bildern und mit Inhaltslinks in den retrodigitalisierten Schulwörterbuchartikeln zu allen Substantivlemmata oder die noch fehlenden Tondateien – wurden bereits im Laufe des Artikels angedeutet.¹⁶ Eine der wichtigsten Entscheidungen ist jedoch, wie die Weiterentwicklung erfolgen könnte, weil daraus unterschiedliche Fragestellungen resultieren: *Welche Module wurden für das Lernerwörterbuch von Anfang an geplant und bisher nicht oder nur teilweise realisiert?* oder: *Welche aktuellen Herausforderungen stellen sich für das Projekt, seitdem die Konzeption und die Datenbank entstanden sind?* Beide Fragen können an dieser Stelle nur ansatzweise beantwortet werden.

Die erste Möglichkeit der Weiterentwicklung ist dementsprechend, die nicht oder nur teilweise realisierten Module wie das Schulwörterbuch-Modul zu vervollständigen, da allein für die Substantivartikel noch ca. 600 didaktisch motivierte Schulwörterbuchartikel zu erstellen sind.¹⁷ Wenn eine nächste Printpublikation ausbleibt, werden nicht nur die Äquivalentangaben zu den Lemmazeichen anderer Wortarten online veröffentlicht, sondern auch die Kollokatoren/Kollokationen sowie die Kombinationen dazu.

Bei der zweiten Möglichkeit wären z. B. folgende Herausforderungen zu nennen: Aktualisierung der Daten durch DWDS-Wortprofil und die korpusgesteuerte Ermittlung bzw. Integration größerer typischer Einheiten sowohl auf der

¹⁶ Folgende erste Herausforderungen sind hier noch zu nennen: Alle Umtexte müssten aktualisiert, dem Onlineformat angepasst, schließlich mit einem entsprechenden Webdesign im Portal integriert werden. Neuentwickelte Benutzungstutorials könnten den Benutzer auf dem Weg zur erfolgreichen Konsultation unterstützen und damit die Lernerautonomie sowie den Lerneffekt erhöhen.

¹⁷ Auch die Schulwörterbuchartikel zu den Verb-, Adjektiv- und Adverb-lemmata müssten im Portal retrodigitalisiert werden, wie es z. B. beim Verb *backen* oder *baden* bereits geschehen ist.

Datenmodellierungs- als auch auf der Präsentationsebene.¹⁸ Letztere bedeutet sowohl eine technische als auch lexikographische Herausforderung mit viel Arbeitsaufwand. Auch eine ein- und/oder mehrsprachige Version des Portals, d.h. ohne Zielsprache oder mit einer anderen Zielsprache als Ungarisch, könnte in Erwägung gezogen werden.

Zuletzt könnte – im Sinne des zum Titel gewählten Terminus *Cross-Media-Publishing* – über die Integration in ein größeres, ungarisch- oder deutschsprachiges Portal sowie über eine App-Version nachgedacht werden.

Welcher Weg der Datenbank des Lernerwörterbuchportals E-KOLLEX DAF in der Zukunft beschieden ist, werden letztendlich die Community und die Ressourcen entscheiden.

References

- Belica, C. (1995): Statistische Kollokationsanalyse und -clustering. Korpuslinguistische Analyseverfahren. Mannheim. <http://corpora.ids-mannheim.de/> (letzter Zugriff: 23-03-2022).
- DW: Deutscher Wortschatz. <https://wortschatz.uni-leipzig.de/de> (letzter Zugriff: 23-03-2022).
- DWDS: Digitales Wörterbuch der Deutschen Sprache. Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart. <https://www.dwds.de> (letzter Zugriff: 23-03-2022).
- E-KOLLEX DAF: Elektronisches Kollokationslexikon Deutsch als Fremdsprache. <http://kollex.hu/szotar/> (letzter Zugriff: 23-03-2022).
- elexiko: <https://www.owid.de/docs/elex/start.jsp> (letzter Zugriff: 23-03-2022).
- Engelberg, S./Müller-Spitzer, C./Schmidt, T. (2016): Vernetzungs- und Zugriffsstrukturen. In: Klosa, A./Müller-Spitzer C. (Hrsg.) (2016) ↑, 153–195.
- Engelberg, S./Storrer, A. (2016): Typologie von Internetwörterbüchern und -portalen. In: Klosa, A./Müller-Spitzer, C. (Hrsg.) (2016) ↑, 31–63.
- Geyken, A. (2011): Statistische Wortprofile zur schnellen Analyse der Syntagmatik in Textkorpora. In: Abel, A./Zanin, R. (Eds.): Korpora in Lehre und Forschung. Bozen, 115–137.
- Gouws, R. H. (2014): Article Structures: Moving from Printed to e-Dictionaries. In: Lexikos 24, 155–177.
- Hausmann, F. J. (1984): Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortverbindungen. In: Praxis des neusprachlichen Unterrichts 31, 395–406.
- Hollós, Z. (2004): Lernerlexikographie: syntagmatisch. Konzeption für ein deutsch-ungarisches Lernerwörterbuch. Tübingen. (Lexicographica. Series Maior 116)
- Hollós, Z. (2008): Kollokationen und weitere typische Mehrwortverbindungen in der ungarischen Lexikographie. In: Hausmann, F. J. (Hrsg.): Collocations in European lexicography and dictionary research. Kollokation in der europäischen Lexikographie und Wörterbuchforschung. Tübingen, 121–133. (Lexicographica 24)

¹⁸ Dies konnte in der Zeit der Datenselektion und -aufbereitung zwischen 2005 und 2009 mangels eines geeigneten Analysetools nur ansatzweise im Datentyp „KOMBIATION“ (siehe Abb. 1) realisiert werden. Davon gibt es z.Z. in der Datenbank 2261 (siehe Tab. 1).

Document template EURALEX

Hollós, Z. (2016): Korpusbasierte intra- und interlinguale Kollokationen. In: Corpas Pastor, G. (Ed.): Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives (Full papers) - Fraseología computacional y basada en corpus: perspectivas monolingües y multilingües (Trabajos completos). Geneva, Switzerland 2016, 302–315. <http://www.tradulex.com/varia/Europhras2015.pdf> (letzter Zugriff: 23-03-2022).

Hollós, Z. (2018): Datendistribution relativ zum Webdesign. Der erste Prototyp des E-KOLLEX. In: Jesenšek, V./Enčeva, M. (Hrsg.): Wörterbuchstrukturen zwischen Theorie und Praxis. Berlin, 151–171. (Lexicographica. Series Maior 154)

Hollós, Z. (2019): Prototyp eines zweisprachigen Internetwörterbuchs für DAF. In: Schierholz, S. J./Klosa, A.: Internetlexikographie und Sprachvermittlung. Berlin/Boston, 65–86. (Lexicographica 34)

HOLL-SULI (2001): Hollós, Z.: Német-magyar Suliszótár. Második, javított és bővített kiadás. Szeged.

Klosa, A. (2007): Korpusgestützte Lexikographie: besser, schneller, umfangreicher. In: Kallmeyer, W./Zifonun, G. (Hrsg.): Sprachkorpora – Dantemengen und Erkenntnisfortschritt. Berlin/New York, 105–122. (Jahrbuch des Instituts für Deutsche Sprache 2006)

Klosa, A./Müller-Spitzer, C. (2016) (Hrsg.): Internetlexikografie. Ein Kompendium. Berlin/Boston.

Lemberg, I. (2001): Aspekte der Online-Lexikographie für wissenschaftliche Wörterbücher. In: Lemberg, I./Schröder, B./Storrer, A. (Hrsg.): Chancen und Perspektiven computergestützter Lexikographie. Tübingen, 71–91.

Müller-Spitzer, C./Koplenig, A./Töpel, A. (2012): Online dictionary use: Key findings from an empirical research project. In: Granger, S./Paquot, M. (Eds.): Electronic Lexicography. Oxford, 425–457.

PONS Bildwörterbuch: www.bildworterbuch.com (letzter Zugriff: 23-03-2022).

Quasthoff, U. (2009): Korpusbasierte Wörterbucharbeit mit den Daten des Projekts *Deutscher Wortschatz*. In: Linguistik online 39, 3/09.

Réder, A. (2006): Kollokationen in der Wortschatzarbeit. Wien.

KOLLEX (2014): Hollós, Z.: SZÓKAPTÁR. Német–magyar SZÓkapcsolatTÁR. Korpusalapú kollokációs tanulószótár. KOLLEX: deutsch-ungarisches KOLLOkationsLEXikon. Korpusbasiertes Wörterbuch der Kollokationen. Deutsch als Fremdsprache. Szeged.

VALBU (2004): Schumacher, H./Kubczak, J./Schmidt R./de Ruiter, V.: VALBU - Valenzwörterbuch deutscher Verben. Tübingen. (Studien zur Deutschen Sprache 31)

WLWF (2010): Wörterbuch zur Lexikographie und Wörterbuchforschung. Dictionary of Lexicography and Dictionary Research. Bd.1. Hrsg. und bearbeitet v. Wiegand, H. E. (et al.). Berlin/New York.

Contact information

Zita Hollós

Károli Gáspár University of the Reformed Church in Hungary (Budapest)

hollós.zita@kre.hu

Acknowledgements

Mein besonderer Dank gilt Uwe Quasthoff, dem Leiter des Projekts *Deutscher Wortschatz*, der die statistisch ermittelten Kookkurrenzlisten dem Projekt zwischen 2005-2009 zur Verfügung gestellt hat. Ich danke auch Cyril Belica am IDS, der den studentischen Hilfskräften bei der zweiten Datenselektion zwischen 2009-2010 den uneingeschränkten Zugriff auf die CCDB ermöglicht hat. Ich möchte auch dem *Ungarischen Nationalfonds für Wissenschaft und Forschung* „OTKA“ (2005-2008) und der *Alexander von Humboldt Stiftung* (2009, 2012, 2013, 2016) meinen Dank aussprechen, die sowohl das Forschungsvorhaben als auch die Realisierung des Projekts finanziell maßgeblich unterstützt haben.