

Carolina Flinz/Laura Giacomini/Weronika Szemińska

TERMIKNOWLEDGE: EIN EINBLICK IN DIE DATENBESCHAFFUNG UND DATENAUFBEREITUNG EINES ONLINE- FACHWÖRTERBUCHS ZUM THEMA COVID-19

Keywords Lexicographic process; LSP dictionary; online dictionary

Die *TermiKnowledge Multilingual Knowledge Base* ist ein Projekt zur Erstellung eines mehrsprachigen korpusbasierten Online-Fachwörterbuchs zum COVID-19-Diskurs, welches sich aktuell im Aufbau befindet (vgl. Storrer 2001).

Vier Universitäten (Universität Warschau, Karls-Universität in Prag, Universität Heidelberg, Universität Mailand) haben sich innerhalb eines 4EU+ Educational Project mit dem Titel „Knowledge through Terminology – From Multilingual Data to Domain-specific Knowledge via Terminological Resources“, abgekürzt *TermiKnowledge*, das Ziel gesetzt, Studierenden von BA- und MA-Studiengängen (Sprach- und Translationswissenschaft sowie Terminologie) Basiskompetenzen zur korpusgestützten und korpusbasierten Fachlexikographie (vgl. Lemnitzer/Zinsmeister 2015, S. 34–37) zu vermitteln, die sie dann zur Erstellung des anvisierten Internetfachwörterbuches (zur Internetlexikographie vgl. Klosa/Müller-Spitzer 2016) unter Betreuung von Fachexpertinnen und Fachexperten einsetzen können.¹ Das Wörterbuch richtet sich an Studierende der Sprach- und Translationswissenschaft, der Terminologie sowie an sonstige Interessierte an diesem Fachdiskurs.

Die Phasen des lexikographischen Prozesses sind von einem Zusammenfließen und Überlappen der Phasen wie in einem Kreis charakterisiert (vgl. Klosa 2013): In der Vorbereitungsphase wurde das Wörterbuch inhaltlich konzipiert und computertechnische Möglichkeiten wurden exploriert. Nach Abwägung unterschiedlicher Möglichkeiten wurde ein Mediawiki-System ausgewählt (vgl. dazu auch Flinz 2018).

In der Phase der Datenbeschaffung wurden die Quellen für die Wörterbuchbasis zusammengestellt. Es handelte sich um *ad hoc* erstellte Vergleichskorpora in deutscher, englischer, italienischer, polnischer und tschechischer Sprache, die aus unterschiedlichen Textsorten kompiliert wurden: normativen Texten (u.a. rechtlichen und medizinischen Leitlinien), wissenschaftlichen Texten (Aufsätzen), Presse-Texten aus allgemeinen Zeitungen sowie Online-Kommentaren zu Presse-Texten. Die Korpora wurden mit der Korpusplattform Sketch Engine aufgebaut und analysiert (vgl. Kilgarriff et al. 2004).

¹ Das Team bestand aus 7 Dozierenden und mehr als 30 Studierenden (7–8 pro Universität und 1 TutorIn pro Universität). Die Arbeit fand online statt; der Kurs (6 cfu) wurde dann jedoch im Frühjahr 2022 mit einem abschließenden Präsenzmeeting in Heidelberg beendet. Die Studierenden haben nach einer theoretischen Einführung in unterschiedlichen Gruppenkonstellationen gearbeitet. In der ersten Arbeitsphase (Erstellung der Korpora und der Lemmalisten) wurden die Gruppen je nach Art von Korpora zusammengestellt; in der zweiten Phase (Herausfilterung der mikrostrukturellen Angaben) erfolgte die Zusammenführung der Studierenden in Gruppen nach Sprachen.

In der Phase der Datenaufbereitung wurden in jeder Sprache die provisorischen Lemmakandidatenlisten erstellt. Diese sind aus dem Zusammenspiel von zwei Arbeitsschritten entstanden: Extrahierung von Frequenzlisten (absoluter und relativer Häufigkeit) aus den unterschiedlichen Fokuskorpora und von Keywordlisten (einzelner Keywords und Mehrwortverbindungen) auf der Basis der in Sketch Engine integrierten Referenzkorpora (Webkorpora). Die provisorischen Stichwortlisten (ca. 57 Lemmata) wurden anschließend untereinander verglichen, um Äquivalenzbeziehungen zu identifizieren (vgl. Flinz/Perkuhn 2018; Szemińska/Więch 2019): z.B. *infection rate* (en) – *Infektionsrate* (de) – *tasso di infezione* (it) – *šíření nákazy* (cz) – *wskaźnik zakażeń* (pl).

Die definitive Lemmaliste (38 Stichwörter), die Entsprechungen in allen involvierten Sprachen hat, wurde dann online gestellt (vgl. Abb. 1):²



Abb. 1: Screenshot der Lemmaliste

Anschließend wurden die Listen validiert und die lexikographischen Daten unter Berücksichtigung der mikrostrukturellen Angaben (u. a. des Formkommentars und des semantischen Kommentars) in einer Datenbank eingetragen. Für jedes Stichwort wurden folgende Informationen aus den Korpora herausgefiltert: semantisch nahe Wörter (auch auf der Basis des Kookkurrenzprofils), orthographische Varianten, Synonyme, Definitionen, Beispiele, Kollokationen und Angaben zur *Keyness*. Wenn möglich wurden die Angaben nach Korpus-typ unterschieden, um auch die intralinguale Variation innerhalb unterschiedlicher Textsorten abzubilden (Abb. 2).

² Vgl. https://terminology.mimuw.edu.pl/index.php?title=Main_Page. Es gab auch Termini, die keine Entsprechung in einer Sprache oder in einer bestimmten Domäne hatten: In diesem Fall wurde ein Hinweis dazu gegeben.

Term:Face mask

Inhaltsverzeichnis [Verbergen]

- 1 Related terms
- 2 Normative
 - 2.1 Variants and synonyms
 - 2.2 Examples
 - 2.3 Collocations
 - 2.4 Keyness
 - 2.5 Note
- 3 Research
 - 3.1 Definition/description
 - 3.2 Examples
 - 3.3 Collocations
 - 3.4 Keyness
 - 3.5 Note
- 4 Press
 - 4.1 Variants and synonyms
 - 4.2 Examples
 - 4.3 Collocations
 - 4.4 Keyness
- 5 Comments
 - 5.1 Variants and synonyms
 - 5.2 Description
 - 5.3 Examples
 - 5.4 Collocations
 - 5.5 Keyness

Abb. 2: Screenshot der Struktur des Eintrags face mask

Das mehrsprachige Wörterbuch wurde bereits online gestellt, auch wenn die Umtexte noch im Aufbau sind.

Die Studierenden haben abschließend die eigene praktische lexikographische Arbeit sowie den theoretischen Hintergrund reflektiert und bewertet: Vor- und Nachteile wurden diskutiert. Dieser letzte Arbeitsschritt war für das Team von großer Relevanz, da Konsequenzen für das bereits gestartete neue *Termiknowledge*-Projekt gezogen werden konnten.

Literatur

Flinz C. (2018): Der lexikographische Prozess bei Tourlex (ein deutsch-italienisches Fachwörterbuch zur Tourismussprache) für italienische DaF-Lerner. In: Klosa, A./Storrer, A./Taborek, J. (Hg.): Internetlexikographie und Sprachvermittlung. Jahrbuch Lexicographica. Berlin, S. 9–36.

Flinz, C./Perkuhn, R. (2018): Wortschatz und Kollokationen in ‚Allgemeine Reisebedingungen‘. Eine intralinguale und interlinguale Studie. In: Krek, S. et al. (Hg.): Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Context. Ljubljana, S. 959–967. <https://euralex.org/publications/wortschatz-und-kollokationen-in-allgemeine-reisebedingungen-eine-intralinguale-und-interlinguale-studie-zum-fachsprachlich-lexikographischen-projekt-tourlex/> (last access: 25-03-2022).

Kilgariff, A. et al. (2004): The Sketch Engine. In: Williams, G./Vessier, S. (eds.): Proceedings of the 11th Euralex International Congress, Lorient, France, July 6–10. Bd. 1. Lorient, S. 105–115.

Klosa, A. (2013): The lexicographical process (with special focus on online dictionaries). In: Gouws, R. et al. (eds.): *Dictionaries. An international encyclopedia of lexicography. Supplementary volume: Recent developments with focus on electronic and computational lexicography*. Berlin/Boston, S. 517–524.

Klosa, A./Müller-Spitzer, C. (Hg.) (2016): *Internetlexikographie. Ein Kompendium*. Berlin/New York.

Lemnitzer, L./Zinsmeister, H. (2015): *Korpuslinguistik. Eine Einführung*. Tübingen.

Storrer, A. (2001): Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie. In: Lemberg, I./Schröder, B./Storrer, A. (Hg.): *Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher*. Tübingen, S. 53–69.

Szemińska, W./Więch, A. (2019): eLex2. A prototype electronic dictionary application for legal translators. In: *Terminology* 25 (2), S. 198–221.

Kontaktinformationen

Carolina Flinz

Università degli Studi di Milano
carolina.flinz@unimi.it

Laura Giacomini

Universität Heidelberg
laura.giacomini@iued.uni-heidelberg.de

Weronika Szemińska

Universität Warschau
w.szeminska@uw.edu.pl