

REPRÄSENTATION VON TERMINOLOGISCHEN VARIANTEN IN RELATIONALEN TERMBANKEN

Laura Giacomini
Universität Heidelberg / Universität Hildesheim

Abstract

Thema des Beitrags ist die Erstellung relationaler Termbanken im Fachübersetzungsunterricht. Inhaltlich und methodisch nimmt der Beitrag Bezug auf Terminologiekurse am Institut für Übersetzen und Dolmetschen der Universität Heidelberg. Der Fokus liegt dabei auf der Repräsentation von Termvarianten und deren Verbindung mit domänenspezifischen Konzepten. Das korpusbasierte Verfahren wird anhand exemplarischer Einträge zu deutschen und italienischen Termini aus dem Fachgebiet der Immunologie dargestellt und umfasst folgende Schritte: Erstellung fachsprachlicher Korpora, Termextraktion, bottom-up Wissensmodellierung, konzeptioneller und logischer Termbankentwurf, Erstellung von Einträgen. Ziel des Beitrags ist es, eine Methode vorzustellen, die es auf Basis von Korpora muttersprachlichen und nicht-muttersprachlichen Übersetzungsstudierenden ermöglicht, terminologische Variation in neuen Fachgebieten zu untersuchen und sie in einer terminografischen Ressource kohärent zu beschreiben.

Keywords: Terminologie; Termbank; relationale Datenbank; terminologische Variation; Wissensmodellierung

Abstract

The topic of the article is the creation of relational termbases in the teaching of technical translation. The article refers to courses in terminology offered at the Institute of Translation and Interpreting at Heidelberg University. The focus is on the representation of term variants and their connection with domain-specific concepts. The corpus-based method is presented on the basis of exemplary entries on German and Italian terms from the field of immunology and comprises the following steps: creation of specialised language corpora, term extraction, bottom-up knowledge modelling, conceptual and logical termbase design, creation of entries. The aim of this contribution is to present a method which enables native and non-native translation students to investigate terminological variation in new subject areas and to describe it coherently in a terminographic resource.

Keywords: terminology; termbank; relational database; term variation; knowledge modeling

1. Einleitung

Ziel des Beitrags ist es, eine Methode vorzustellen, die es auf Basis von Korpora muttersprachlichen und nicht-muttersprachlichen Übersetzungsstudierenden ermöglicht, terminologische Variation in neuen Fachgebieten zu untersuchen und sie in einer terminografischen Ressource systematisch zu beschreiben.

Das korpusbasierte Verfahren wird anhand exemplarischer Einträge zu deutschen und italienischen Termini dargestellt. Die hier vorgestellten Daten stammen aus der Fachsprache der Immunologie, die in zwei aufeinanderfolgenden Lehrveranstaltungen analysiert wurde. Die Methode, die vor dem ersten Kurs entwickelt und während der beiden Kurse verfeinert wurde, ist auf die Bedürfnisse von Muttersprachlern und Nicht-Muttersprachlern des Italienischen und Deutschen abgestimmt. Dies gewährleistet eine größere Flexibilität in Bezug auf die Zielgruppe. Das Verfahren umfasst eine Reihe von Schritten, die im Folgenden aufgeführt und kurz beschrieben werden sollen.

Dieser Beitrag beschäftigt sich mit der Erstellung exemplarischer bzw. prototypischer relationaler Termbanken im Fachübersetzungsunterricht. Inhaltlich und methodisch nimmt der Beitrag Bezug auf Terminologiekurse für Master-Studierende des Instituts für Übersetzen und Dolmetschen der Universität Heidelberg. Der Fokus liegt dabei auf der Repräsentation von Termvarianten und deren Verbindung mit domänenspezifischen Konzepten. Dieser Schwerpunkt ermöglicht es, den technischen Aspekt der Planung und Erstellung einer für Übersetzer konzipierten Datenbank mit einem heute besonders wichtigen theoretischen Aspekt der Fachkommunikation zu verbinden, nämlich dem Phänomen der terminologischen Variation. Kapitel 2 führt die Ziele des Beitrags und die bei der Erstellung der Terminologiedatenbanken angewandte Methode ein. In Kapitel 3 werden Termvarianten als zentraler Datentyp aus theoretischer und anwendungsorientierter Sicht dargestellt. Kapitel 4 wird sich speziell mit dem Design und der Implementierung einer Datenbank in Xampp (s. Kap. 4.2) befassen. Schließlich werden in Kapitel 5 die Vorteile und das Potenzial der Methode zusammengefasst.

2. Hintergründe und Methode

Eine der größten Herausforderungen bei der Vermittlung von Kompetenzen im Bereich Korpusarbeit in Studiengängen für Übersetzer ist die zeitliche Begrenzung der einzelnen Lehrveranstaltungen. Der Erwerb von Wissen im Bereich der Korpuslinguistik beschränkt sich daher in der Regel auf theoretische Grundlagen und Telexperimente, die eventuell in BA- und MA-Arbeiten vertieft werden können. Für künftige Übersetzer wäre es jedoch sinnvoll, selbst im Rahmen eines Kurses umfassendere Kompetenzen zu erwerben und zu erfahren, wie man terminografische Prozesse plant und umsetzt, die in der heutigen Arbeitswelt von Bedeutung sind. Es ist demnach notwendig, Unterrichtsziele und -methoden zu entwickeln, die zu greifbaren und reproduzierbaren Ergebnissen in den Kursen führen können, und zwar durch Experimente, die zwar in ihrem Umfang klein, aber in ihrer Konzeption vollständig sind. Die Arbeit mit Korpora erfordert sowohl linguistisches als auch technisches Wissen, das beispielsweise von der Annotation eines Korpus bis zur Darstellung der aus Korpora extrahierten Daten in geeigneten Datenbanken reicht. Die Erstellung von Datenbanken wiederum ist ein Thema, das eng mit zahlreichen technisch-formalen Aspekten verbunden ist, z.B. gängigen Datenrepräsentationsformaten.

2.1 Erstellung fachsprachlicher Korpora

Die Wahl eines bestimmten Fachgebiets ist ein wichtiger Bestandteil der geplanten Terminologiarbeit. Im Rahmen einer Lehrveranstaltung ist es von Vorteil, sich für ein relativ begrenztes Fachgebiet bzw. Teilgebiet mit einer spezifischen Terminologie zu entscheiden, für das ausreichend Textmaterial verfügbar ist.

Für jede im Kurs festgelegte Sprache wird ein Korpus von Fachtexten zu dem gewählten Fachgebiet mithilfe der Funktionen von *Sketch Engine* (vgl. Kilgarriff et al. 2014) erstellt, in dem vorhandene Fachtexte, z.B. aus Fachübersetzungskursen, verwendet oder Fachtexte online gesammelt werden. In beiden Fällen wird die Textsammlung automatisch vorverarbeitet und annotiert, sodass sie später mit den Werkzeugen von *Sketch Engine* abgefragt werden kann. Der Schwerpunkt dieser Arbeitsphase liegt auf der Erstellung repräsentativer und textsortenspezifisch homogener Korpora. Es muss also im Vorfeld die zu analysierende kommunikative Ebene bestimmt werden, die wiederum vom vorgesehenen Endnutzer/der vorgesehenen Endnutzerin der Termbank abhängig ist. Jedes Korpus enthält etwa 1 Million Wörter, eine Größe, die in der Regel sowohl hinsichtlich der verfügbaren Zeit als auch des beabsichtigten Zwecks angemessen ist. Die Möglichkeit der Erstellung

von Parallelkorpora bleibt bestehen, wenn Translation Memories verfügbar sind (z.B. tmx-Dateien aus früheren Übersetzungsprojekten). Für die Erfassung von Daten für die vorgesehene Termbank ist dies jedoch nicht unbedingt erforderlich.

2.2 Term- und Variantenextraktion

Der Erstellung von Korpora folgt die Extraktion von Termini mithilfe des Keywords-Tools. Das Tool extrahiert automatisch aus einem Korpus Listen von Termkandidaten, die auf der Grundlage eines Vergleichs zwischen dem Fachkorpus und einem Referenzkorpus, in der Regel ein gemeinsprachliches Korpus, ermittelt wurden. Die extrahierten Kandidaten sind sowohl einfache als auch komplexe Termini, darunter Fachkollokationen und Mehrworttermini.

Die Validierung der Kandidaten ist eine zentrale Phase der Arbeit. Hier gilt es u.a., fehlende Fachkenntnisse durch die Dokumentation in lexikografischen und terminografischen Quellen und Paralleltexten auszugleichen. Ziel dieser Phase ist es, Termini und ihre Varianten zu sammeln, um sie als Lemmata in die Datenbank aufzunehmen. Der *main term* (i.S.v. Vorzugsbenennung), gegenüber dem alle Synonyme als Varianten zu betrachten sind, wird entweder auf qualitativer Basis (es handelt sich um die normierte Form) oder auf quantitativer Basis (es handelt sich um die häufigste Form im Korpus) festgelegt.

2.3 Wissensmodellierung

Die Terminologiedatenbank enthält nicht nur lexikalische Daten (Termini und Varianten), sondern auch konzeptuelle Daten, die mit ersteren verknüpft sind. Das Hauptziel besteht darin, die Beschreibung der Terminologie des gewählten Bereichs kohärent und systematisch zu gestalten, indem die korrekte Zuordnung jedes Terminus zu einem oder mehreren Konzepten gewährleistet wird. Auf diese Weise wird ein einfaches konzeptuelles Wissensmodell erstellt, das eine bestimmte Anzahl an Konzepten beinhaltet. Die Granularität dieser Art von Daten wird auf der Grundlage verschiedener Kriterien gewählt, z.B. der konzeptuellen Merkmale der Domäne, des Umfangs des terminologischen Datenbestandes und der verfügbaren Zeit. Es ist daher möglich, sehr detaillierte oder allgemeinere konzeptuelle Netze zu erstellen, beispielsweise durch die Ermittlung von Makro-Konzepten, die man informell als ‚Themen‘ bezeichnen könnte (z.B. GESUNDHEITSWESEN, KRANKHEITEN, MEDIKAMENTE, THERAPIEN).

Je detaillierter die konzeptuelle Datenbasis, desto aufwendiger ist es, Relationen zwischen den Konzepten zu ermitteln. Diese Relationen ermöglichen u.a., lexikalische Zusammenhänge zwischen den Termini zu identifizieren. In vielen Fällen kann ein an semantischen Rollen oder an Frames orientierter Ansatz dabei sehr hilfreich sein. Semantische Rollen ermöglichen es, Konzepte mit bestimmten Rollenprofilen in der Domäne zu gruppieren, z.B. AGENT, PATIENT, BENEFICIARY. Frames erlauben es, typische Szenarien innerhalb einer Domäne zu identifizieren (z.B. ZULASSUNG EINES MEDIKAMENTS oder THERAPIE GEGEN EINE VIRUSINFEKTION), mit denen eine typische Konstellation von Konzepten und Begriffen verbunden ist (vgl. Faber 2015; Giacomini 2018, 2019). Konzepte, die ein Frame identifizieren, sind seine Elemente, die durch logische und funktionale Beziehungen miteinander verbunden sind. Einige mögliche Elemente des Frames THERAPIE GEGEN EINE VIRUSINFEKTION könnten z.B. sein: MEDIZINISCHES PERSONAL, MEDIKAMENTE, PATIENTEN. Im Wesentlichen handelt es sich um Konzepte, die in Texten auf unterschiedliche Weise lexikalisiert werden. Auch bei einem frame-orientierten Ansatz kann diejenige Granularität festgelegt werden, die für das Experiment am besten geeignet ist.

Ähnlich komplexeren Wissensmodellen wie Ontologien, geht es auch bei einfacheren Begriffssystemen nicht nur um die Beschreibung von (Klassen von) Objekten – in diesem Fall Konzepten –, sondern auch um deren Relationen. Unabhängig davon, welches konzeptuelle Modell für das jeweilige Experiment am besten geeignet ist, müssen Relationen zwischen Konzepten explizit dargelegt und benannt werden, gegebenenfalls auf informelle Weise. Neben grundlegenden hierarchischen konzeptuellen Relationen wie den Typ-von- und Teil-von-Relationen kann es eine Vielzahl anderer nicht-hierarchischer Relationen geben, z.B. funktionaler oder kausaler Natur (vgl. Arntz et al. 2014; Giacomini 2019).

Die Identifizierung von Konzepten ist ein Verfahren, das mit der Analyse der Texte aus den Korpora beginnt und mit der Beobachtung der extrahierten Termini fortgesetzt wird. Diese können zunächst in breite semantische Felder gruppiert und dann schrittweise eingegrenzt werden. Dieses Bottom-up-Verfahren ist grundlegend, um eine lexikalische Ressource zu schaffen, die das in den Korpora enthaltene Fachvokabular lexikalisch und konzeptuell widerspiegelt. Das Verfahren muss allerdings durch die Dokumentation in externen Quellen (Paralleltexten, lexikografischen und terminografischen Quellen, Handbüchern usw.) unterstützt werden. Während diese Aufgabe im Kontext größerer Forschungsvorhaben idealerweise von einem Top-Down-Verfahren begleitet werden sollte – d.h. ausgehend von bestehenden konzeptuellen Systemen, z.B. Ontologien –, ist dieser Weg in eng begrenzten Experimenten, wie wir sie hier beschreiben, weniger praktikabel, da er komplex und zeitaufwändig ist.

2.4 Entwurf eines konzeptionellen, eines logischen und eines physischen Datenmodells für die relationale Termbank

Eine relationale Datenbank ist z. Z. neben dokumentorientierten XML-Datenbanken ein gebräuchliches Mittel zur Repräsentation lexikografischer bzw. terminografischer Daten. In diesem Datenbankmodell werden Daten in Tabellen gespeichert, die Attribute enthalten und miteinander über Relationen verknüpft sind. Daten werden dabei durch ein Datenbankmanagementsystem (DBMS) verwaltet und ausgelesen.

Bevor die Einträge in der Terminologiedatenbank erfasst werden können, müssen die Struktur der Datenbank und die Art der darin enthaltenen Daten jedoch festgelegt werden. Dies hängt im Wesentlichen von den Bedürfnissen des idealen Benutzers/der idealen Benutzerin der geplanten lexikalischen Ressource ab. Die Erstellung einer Terminologiedatenbank erfordert also ein Datenmodell, das die zu beschreibenden und zu verarbeitenden Daten innerhalb einer Domäne sowie deren Relationen untereinander festlegt. Der Zweck der Datenmodellierung ist also die Sortierung und Strukturierung von Daten und man unterscheidet zwischen konzeptioneller, logischer und physischer Modellierung – dabei handelt es sich um Begriffe, die größtenteils aus den Bereichen der Softwaretechnik und der Wirtschaftsinformatik stammen.

Ein konzeptionelles Datenmodell ist ein „Datenmodell, das die globale logische Struktur aller Daten [...] (oder zumindest eines mit einem Datenbanksystem erfassten Teilbereichs) implementierungsunabhängig beschreibt und diese in einer fassbaren und systematischen Form strukturiert darstellt“ (Gabler Wirtschaftslexikon, <https://wirtschaftslexikon.gabler.de>). In unserem Falle spezifiziert das konzeptionelle Datenmodell lexikografische bzw. terminografische Angabetypen und ihre Beziehungen zueinander. Hierfür stehen nützliche Beschreibungsformate, etwa UML-Klassendiagramme oder Entity-Relationship-Diagramme, zur Verfügung.

Diagramme in der Unified Modelling Language (UML) dienen zur grafischen Darstellung von Klassen von Objekten (hier: Typen von lexikalischen bzw. konzeptuellen Daten). Im Falle einer relationalen Terminologiedatenbank entspricht jede Klasse einer bestimmten Tabelle, die wiederum eine Reihe von Attributen enthält (Abbildung 1).

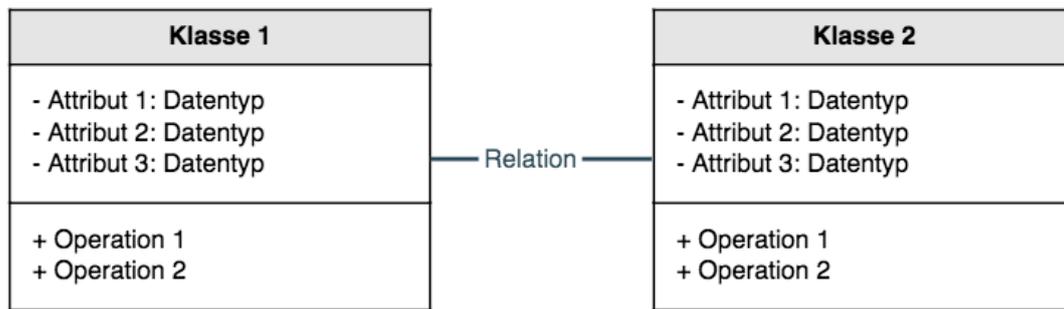


Abbildung 1

Klassen, Attribute, Operationen und Relationen in UML-Diagrammen. Zu jedem Attribut gehört auch ein Datentyp.

Attribute einer Klasse sind nichts anderes als lexikografische oder terminografische Angaben, die in einer bestimmten Tabelle enthalten sind (z.B. grammatische Angabe, Aussprachenangabe, Definition, Beispielangabe, pragmatische Angabe usw.). Die Klassen, grafisch durch Rechtecke dargestellt, sind durch Relationen miteinander verbunden, grafisch durch Linien dargestellt. Während das UML-Modell im Allgemeinen auch die Möglichkeit bietet, Relationen der Generalisierung, Aggregation und Komposition zu identifizieren, gibt es im hier vorgestellten lexikografischen Modell nur Assoziationsrelationen, d.h. jede Tabelle in der Datenbank muss mit mindestens einer anderen Tabelle assoziiert sein. In Abbildung 1 wird auch gezeigt, welche Operationen für eine bestimmte Menge von Attributen in einer Datenbank möglich sein werden.

Bei den Verbindungslinien zwischen Klassen können darüber hinaus Namen für die Relationen eingefügt und sog. Multiplizitäten angegeben werden. Hierbei handelt es sich um Werte, die die Assoziationen in beiden Richtungen ‚quantifizieren‘. Multiplizitäten drücken folgende Informationen aus:

- 0..*: ein Objekt einer Klasse verweist auf eine beliebige Zahl von Objekten einer anderen Klasse.
- 1..*: ein Objekt einer Klasse entspricht mindestens einem, ansonsten aber beliebig vielen Objekten einer anderen Klasse.
- 1: ein Objekt einer Klasse entspricht genau einem Objekt einer anderen Klasse.

Die Verwendung von Multiplizitäten ermöglicht es, frühzeitig über die Zusammenhänge zwischen den verschiedenen lexikografischen Angaben eines Datenbankeintrags nachzudenken und fundierte Entscheidungen zu treffen. In dieser Phase wird zum Beispiel festgelegt, ob es für ein Lemma mehr als eine Definition geben darf oder ob Beispiele obligatorisch sind.

Im nächsten Schritt wird ein logisches Datenmodell entworfen, d.h. ein „auf die spätere Implementierung ausgerichtetes Datenmodell, das die Daten für den späteren Einsatz bereits vorstrukturiert“ (Gabler Wirtschaftslexikon, <https://wirtschaftslexikon.gabler.de>). Dabei wird das konzeptionelle Datenmodell für ein bestimmtes Datenformat und das damit verbundene Datenbanksystem spezifiziert. Bei relationalen Datenbanken geht es meistens um eine Strukturierung der Datenelemente, z.B. die Bestimmung der einzelnen attributbezogenen Datentypen. So können beispielsweise Art und Länge von lemma-, definitions- und kollokationsidentifizierenden Strings bestimmt werden. Primärschlüssel zu jeder Klasse (Tabelle) können während dieser Phase bestimmt werden.

Der Entwurf des physischen Datenmodells schließt diesen Prozess ab und definiert die endgültige Form der Datenbank. Klassen- und Attributnamen werden auf reservierte Wörter geprüft und gegebenenfalls angepasst, damit sie für das DBMS lesbar sind. Primärschlüssel, falls noch vorhanden, Fremdschlüssel und Einschränkungen (*constraints*) für die Schlüssel werden ebenfalls eingeführt.

Sobald alle Schlüssel festgelegt worden sind und somit die (noch leeren) Tabellen der relationalen Datenbank im DBMS erstellt werden, kann die Datenbank normalisiert werden. Die Normalisierung impliziert die Anwendung bestimmter Regeln zur sinnvollen Aufteilung von Attributen in mehreren Tabellen. Ziel dabei ist die grundsätzliche Vermeidung von Redundanzen und Anomalien (insbesondere funktioneller und transitiver Abhängigkeiten), die das Datenbankschema unnötig erschweren würden.

In Kapitel 4 wird auch dargelegt, wie bei der Terminologearbeit die Erstellung einer vorläufigen Gesamttabelle die Modellierungsphase unterstützen kann, weil sie auf der Grundlage einer ersten Datensammlung die Diskussion über Ziele, mögliche Probleme und geeignete Strategien fördert. Diese Gesamttabelle, die später in mehrere Tabellen gespalten werden soll, dient der Ergänzung der konzeptionellen, logischen und physischen Modelle, kann aber diese nicht ersetzen.

2.5 Erstellung von Termbank-Einträgen

Sobald die Modellierung der Daten abgeschlossen ist, erfolgt die Erstellung der einzelnen Tabellen der Datenbank, die mit den zuvor erhobenen und validierten lexikalischen und konzeptuellen Daten gefüllt werden. Ein Eintrag im lexikografischen und terminografischen Sinne entspricht der Vereinigung der Datensätze, die in verschiedenen, durch Fremdschlüssel miteinander verbundenen Tabellen und ausgehend von einer Haupttabelle identifiziert werden können. Die Tabelle, die wir als *Haupttabelle* definieren, enthält alle Lemmata, also die Daten, die typischerweise im Mittelpunkt einer lexikalischen Ressource stehen. In dieser Phase verfügen wir bereits über Konzepte, Termini und Varianten, die in die Tabellen als Lemmata oder als sonstige Angaben (z.B. Synonyme, Kollokationen, Verweise) eingefügt werden können. Nun müssen aber auch andere Arten von Angaben, die im Modell der Terminologiedatenbank vorgesehen sind, je nach Bedarf aus den Korpora gewonnen werden.

Aus Korpora können nicht nur Daten über die Kollokationen von Lemmata (in *Sketch Engine* durch das *Word Sketch Tool*) und Beispiele (durch Konkordanzen) gewonnen werden, sondern auch Daten über semantische Relationen. Diese können durch die Analyse bestimmter Textmuster erhoben werden, die für eine Textdomäne typisch sein können (z.B. Synonymie in Mustern wie *ein X ist auch als Y bekannt* oder Hyperonymie in Mustern wie *X ist eine Art von Y*).

Der Prozess der Akquisition von Äquivalenten hängt von der Art der verwendeten Korpora ab. Im Falle von Parallelkorpora können Äquivalentkandidaten extrahiert und validiert werden. Stehen hingegen nur einsprachige Korpora zur Verfügung, so können diese in der Regel nur zur Validierung von Äquivalenten genutzt werden, die in Paralleltexten und durch den Austausch mit Fachexperten gesucht werden. Im Mittelpunkt dieses Beitrags stehen der vierte und der fünfte Schritt (Kap. 2.4 und 2.5), die sich speziell auf die Erstellung der terminologischen Datenbank beziehen.

3. Terminologische Variation

Das Thema der terminologischen Variation in der Terminologieforschung hat in den letzten zwanzig Jahren grundlegende Bedeutung erlangt. Zahlreiche Forscher:innen haben das natürliche Vorkommen synonymischer Varianten im fachspezifischen Lexikon vieler Disziplinen zunehmend hervorgehoben und dieses Phänomen kann somit bei der Terminologie- und der Korpuserarbeit nicht mehr ignoriert werden. Eine strikte, standardisierende und eher an der traditionellen Terminologielehre orientierte Perspektive von Variation als unerwünschte Erscheinung ist heutzutage aus vielen Gründen nicht mehr praktikabel. Die heutige Forschungsrichtung wurde von teilweise sehr unterschiedlichen Anliegen angetrieben, etwa der Suche nach einem geeigneten Modell für die Extraktion von Termini in der

Computerlinguistik bzw. für die Repräsentation von Termini in lexikografischen und terminografischen Ressourcen (vgl. Giacomini 2021). Die verfügbaren Klassifikationsmodelle für terminologische Varianten sind entsprechend heterogen, funktional ausgerichtet und weisen meistens nur partielle Überschneidungen auf.

In diesem Beitrag wird der Fokus auf nicht-diasystemischer Variation liegen, d.h. auf Varianten, die grundsätzlich keine diachronischen, diatopischen, diamesischen, diastratischen oder diaphasischen Unterschiede aufweisen und innerhalb desselben Textes vorkommen können. Diese Art der Variation ist in der naturwissenschaftlichen und technischen Fachsprache weit verbreitet. Registerunterschiede können dabei vorkommen, meistens geht es allerdings um einfache Synonyme bzw. Quasi-Synonyme, die in der gleichen Situation und unter den gleichen Kommunikationspartner:innen verwendet werden. Diese Synonyme stellen für Übersetzer auch deshalb eine Herausforderung dar, weil sie in der Regel nicht von lexikografischen bzw. terminografischen Ressourcen verzeichnet werden. Ihre Identifizierung und Übersetzung erfordert oft zeitaufwändige Dokumentationsarbeit, insbesondere durch die Recherche in Paralleltexten. Dies geschieht nicht nur bei der Übersetzung in die Fremdsprache, sondern auch in die Muttersprache, sobald die fachlichen Kompetenzen des Übersetzers nicht ausreichen, um Fragen wie diese zu beantworten: Ist eine bestimmte terminologische Variante in Gebrauch? Ist sie in einem bestimmten Fachgebiet, in einer bestimmten Textsorte, auf einer bestimmten Ebene der Fachkommunikation verwendbar?

In diesem Artikel beziehen wir uns auf die in Giacomini (2019) entwickelte Variantentypologie, die für jedes Paar aus Hauptterminus (i.S.v. Vorzugsbenennung) und Variante die Angabe der Werte von drei Variationstypen vorsieht, nämlich der orthografischen Variation (OV), der syntaktischen Variation (SV) und der lexikalisch-morphologischen Variation (MV). Angaben zur lexikalisch-morphologischen Variation orientieren sich an der Übereinstimmung oder Nichtübereinstimmung von lexikalischen Morphemen innerhalb der gegenübergestellten Termini.

Während die ersten beiden Typen einfach vorhanden (+) oder nicht-vorhanden (-) sein können, ist es im Fall der lexikalisch-morphologischen Variation möglich, den zusätzlichen Wert des partiellen Auftretens (/) der Variation anzugeben. Im Folgenden werden zwei Beispiele für Varianten und Variationstypen aus dem Bereich der Immunologie genannt:

(a) Reproduktionszahl – R-Wert	MV+	SV-	OV-
(b) Linfocita T – cellula T	MV/	SV-	OV-
(c) COVID-19 – Covid-19	MV-	SV-	OV+
(d) WHO – World Health Organization	MV+	SV+	OV-

In Beispiel (a) finden wir ein Terminuspaar ohne syntaktische oder orthografische Variation, jedoch mit vollständiger lexikalisch-morphologischer Variation. Bei Beispiel (b) ist eine partielle lexikalisch-morphologischer Variation vorhanden: Das Element *T* tritt in beiden Formen auf, während *linfocita* und *cellula* andere lexikalische Morpheme aufweisen. Im dritten Beispiel (c) sind keine morphologischen oder syntaktischen Unterschiede zwischen Hauptterminus und Variante zu erkennen. Die Variation ist lediglich orthografischer Natur und betrifft die Groß- bzw. Kleinschreibung in den zwei Termini. Das letzte Terminuspaar (d) zeigt hingegen ein Beispiel für syntaktische Variation, das typischerweise bei der Auflösung eines Initialismus oder eines Akronyms auftritt. Gleichzeitig ist eine vollständige lexikalisch-morphologische Variation zu sehen.

Bei dieser Auffassung von Variation finden sich größtenteils Fachkollokationen, deren Varianten auf die relative Modifizierbarkeit und Ersetzbarkeit der Kollokationsbestandteile zurückzuführen sind. Diese Varianten treten als Koreferenten in einem Text oder als Alternativen in verschiedenen kommunikativen Kontexten, Genres und Texttypen auf.

4. Die Erstellung einer relationalen Termbank der Immunologie

In diesem Kapitel wird der Schwerpunkt auf den Entwurf von ImmunoTerm, der Terminologiedatenbank für das Fachgebiet der Immunologie, und die Erfassung ihrer Einträge gelegt. Der Bereich der Immunologie wurde gewählt, um den Kurs über Datenbanken thematisch mit einer parallelen Fachübersetzungsübung Deutsch-Italienisch zu begleiten, die dasselbe Thema zum Inhalt hatte. Aufgrund der aktuellen Pandemie ist die Immunologie zu einem bekannten Thema geworden. Dies hat möglicherweise dazu beigetragen, dass die Terminologie des Fachgebiets auf der untersten kommunikativen Ebene der vertikalen Gliederung der medizinischen Fachsprache, d.h. auf der Ebene der Kommunikation mit Patienten oder interessierten Nichtexperten, ausgeweitet wurde. Im Allgemeinen weist die Fachlexik der Immunologie in den untersuchten Sprachen die typischen Merkmale der medizinischen Fachsprache auf, u.a. die starke Präsenz von lexikalischen Formen neoklassischen Ursprungs und Fachphrasemen, sowie die ausgeprägte Dualität zwischen Sprache der Wissenschaft und Sprache der Popularisierung (vgl. Serianni 2005; Pera / Schmiedebach 2007; Cortelazzo 2009; Busch / Spranz-Fogasy 2015).

4.1 Termbankentwurf

Das hier beschriebene Beispiel einer terminologischen Datenbank basiert auf einer zweisprachigen Ressource für Italienisch und Deutsch, die bidirektional und bifunktional angelegt ist. Dies setzt eine strukturelle Gleichheit der Einträge in den beiden Sprachen voraus.

Die ImmunoTerm Datenbank soll als Grundlage für ein zweisprachiges Fachwörterbuch dienen. Der ideale Benutzer/die ideale Benutzerin ist der Fachübersetzer/die Fachübersetzerin, dessen/deren Muttersprache Deutsch oder Italienisch ist und der Texte typischerweise in die L1, manchmal aber auch in die L2 oder Fremdsprache übersetzt. Dieser Benutzer/diese Benutzerin verfügt über fortgeschrittene Sprachkenntnisse in beiden Sprachen und über durchschnittliche Fachkenntnisse: Er/Sie kann das Thema und die Fachsprache weitgehend verstehen, jedoch nicht auf demselben Niveau wie ein Fachexperte. Die in der Datenbank enthaltene Terminologie ist in Bezug auf das kommunikative Niveau übergreifend, d.h. sie umfasst sowohl hochspezialisierten Wortschatz als auch das für popularisierende Textsorten typische Fachvokabular.

Lexikalische und konzeptuelle Daten (in unserem Fall: lexikografische Angaben) sind in Tabelle 1 aufgelistet und werden nach dem Mikrostrukturmodell von Wiegand (vgl. Bergenholtz et al. 2008) in Formkommentar und semantischen Kommentar unterteilt:

Formkommentar:	Semantischer Kommentar:	
Terminus (mit TermID)	Kommunikative Ebene/Register	Verweis
Sprache	Definition	Abbildung
IPA	Definitionsquelle	KONZEPT
Wortklasse	Variante	
Genus	Äquivalent	
Genitiv	Beispiel	
Plural	Beispielquelle	

Tabelle 2
Liste der Datentypen in der Termbank.

Im semantischen Kommentar geben wir auch an, welches Konzept (in Großbuchstaben) durch einen Terminus ausgedrückt wird. Bei den lexikografischen Angaben, die speziell für unser Experiment erstellt wurden, fällt auf, dass es kein Attribut gibt, das der Angabe der Kollokationen des Lemmas

dient. Diese Wahl lässt sich damit begründen, dass wir beabsichtigen, Kollokationen als Lemmata (Attribut „Terminus“) in gleicher Weise zu behandeln wie einfache Termini. Die Beziehung zwischen einem einfachen (Fach-)Wort und seinen Kollokationen wird durch Querverweisangaben (Attribut „Verweis“) ausgedrückt. Diese Angabetypen sind, wie in Kapitel 2.4 erläutert, in einer einzelnen, allgemeinen Tabelle zusammengefasst, die alle dargestellten Attribute enthält und die für eine erste Einschätzung der Datenabdeckung nützlich ist.

Bei der Modellierung der Daten wird auch der Datentyp jedes Attributs festgelegt, der in unserem Fall VARCHAR(n) ist, d.h. eine Zeichenkette mit einer variablen Länge. Die einzige Ausnahme bildet das Attribut TermID, der numerische Bezeichner jedes Terminus, der vom Typ INTEGER(n) ist, d.h. eine Ganzzahl.

Jede Klasse im UML-Diagramm entspricht einer Tabelle in der Datenbank. Insgesamt besteht die relationale Datenbank aus 8 Tabellen. Die Struktur der normalisierten Tabellen ist nachstehend aufgeführt (jedem Tabellennamen geht ein „t_“ vor):

(1) t_haupttabelle:

<u>TermID</u>	Terminus	<u>Sprache</u>	IPA	Wort- klasse	Genus	Genitiv	Plural	Komm. Ebene	Konzept	Abb.
---------------	----------	----------------	-----	-----------------	-------	---------	--------	----------------	---------	------

(2) t_definitionen	<u>TermID</u>	<u>Definition</u>	
(3) t_definitionsquellen	<u>Definition</u>	<u>Quelle</u>	
(4) t_varianten	<u>TermID</u>	<u>Variante</u>	Variationstyp
(5) t_äquivalente	<u>TermID</u>	<u>Äquivalent</u>	
(6) t_beispiele	<u>TermID</u>	<u>Beispiel</u>	
(7) t_beispielquellen	<u>Beispiel</u>	<u>Quelle</u>	
(8) t_verweise	<u>TermID</u>	<u>Verweis</u>	

Primärschlüssel wurden durch Unterstreichen des Namens des entsprechenden Attributs gekennzeichnet, z.B. TermID in t_varianten. In einigen Fällen gibt es zusammengesetzte Schlüssel, die mehr als einem Attribut entsprechen, z.B. TermID und Sprache in t_haupttabelle. Diese Lösung ist nützlich, wenn ein Datensatz (d.h. eine Zeile in einer Tabelle) nicht durch ein einziges Attribut eindeutig identifiziert werden kann. Dies ist bei der t_haupttabelle der Fall, bei der ein deutscher Terminus und ein italienischer Terminus potenziell homonym sein könnten (siehe Eigennamen, Abkürzungen, Fremdwörter usw.) und daher nur anhand der Kombination von Terminus und Sprache zu unterscheiden sind.

Das UML-Diagramm in Abbildung 2 veranschaulicht die Struktur der Terminologiedatenbank mit sämtlichen Klassen, Attributen, Datentypen und Primärschlüsseln.

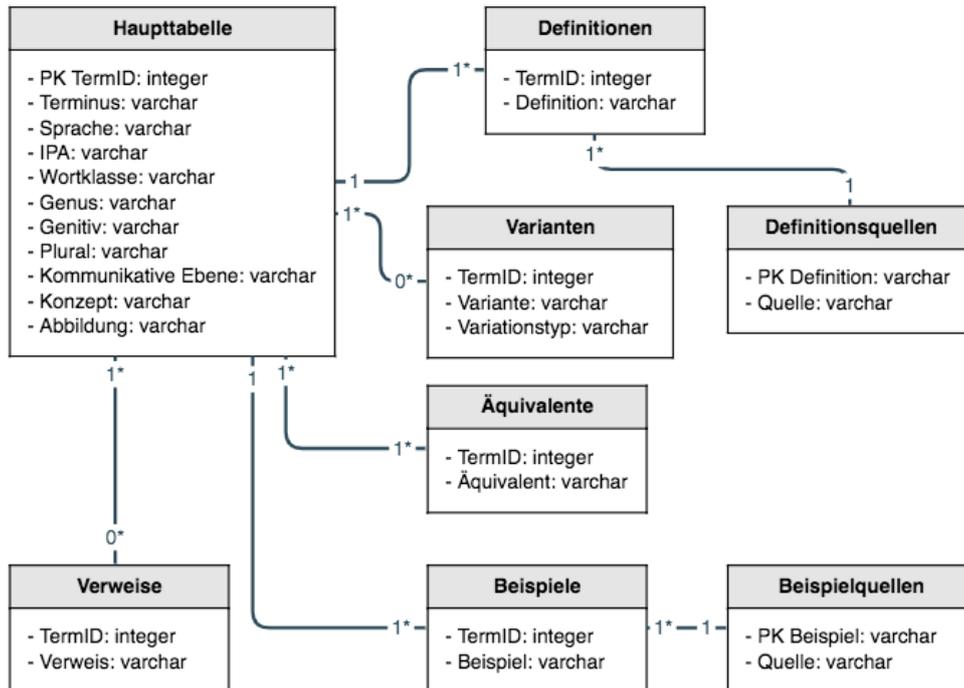


Abbildung 2

UML-Diagramm mit dem physischen Datenmodell, aus dem Klassen, Attribute, Datentypen und Relationen mit Multiplizitäten ersichtlich sind. PK: Primary Key (Primärschlüssel).

Es werden auch Multiplizitäten angezeigt, die die Art der Relationen zwischen den Klassen angeben. Aus dem UML-Diagramm wird nun deutlich, dass eine Terminologiedatenbank geplant ist, in der das Lemma mindestens eine Definition, ein Äquivalent und ein Beispiel haben muss. Das Lemma muss allerdings nicht unbedingt eine Variante haben oder auf etwas verweisen. Jede Definition und jedes Beispiel hat genau eine Quelle, während dieselbe Quelle mit mehr als einer Definition und mehr als einem Beispiel verbunden sein kann.

Der Übersichtlichkeit halber wurde beschlossen, eine ID nur mit dem Attribut „Terminus“ zu verknüpfen, während in anderen Fällen der Primärschlüssel dem Attribut selbst entspricht.

Die Datenbank enthält zwei Arten von Tabellen: Tabellen mit Primärschlüsseln bzw. Fremdschlüsseln (1, 3, 7) und Verbindungstabellen, bei denen das Schlüsselattribut aus einer Verknüpfung von zwei Schlüsseln besteht (2, 4, 5, 6, 8). Diese verschiedenen Typen spiegeln die in Abbildung 2 gezeigten unterschiedlichen Multiplizitäten wider, d.h. die verschiedenen Relationen zwischen den Tabellen.

Die t_haupttabelle enthält Attribute, mit denen das Lemma (TermID/Terminus) folgende Relationen aufweist:

1:1 (ggf. 1:0), d.h. einem und nur einem Lemma entspricht genau eine Pluralform, eine Genitivform bzw. eine Abbildung;

n:1 (ggf. n:0), d.h. einer bestimmten Anzahl n von Lemmata entspricht genau eine Sprache, eine Wortklasse, eine kommunikative Ebene bzw. ein Konzept.

An diesem Beispiel wird deutlich, dass diese Struktur das Resultat gezielter Entscheidungen ist, z.B. bei der Zuordnung von Termini zu Konzepten. In einer anderen Terminologiedatenbank könnte man sich dafür entscheiden, mehr als ein Konzept für denselben Terminus zuzulassen.

Die Tabellen `t_definitionsquelle` und `t_beispielquelle` haben die gleiche Art von Relation wie `t_haupttabelle`, d.h. die Quelle kann einer oder mehreren Definitionen bzw. einem oder mehreren Beispielen zugeordnet werden. Der Unterschied zur Haupttabelle besteht darin, dass die Quelle nicht ein Attribut des Lemmas, sondern der Definition oder des Beispiels ist. Durch Anwendung der dritten Normalform auf die ursprüngliche Gesamttabelle werden zwei getrennte Tabellen erstellt, in die die Quellen für Definitionen und Beispiele eingefügt werden können.

Verbindungstabellen weisen hingegen eine n:m-Relation zwischen ihren Attributen auf, d.h. eine n-Anzahl von Instanzen eines bestimmten Attributs entspricht einer m-Anzahl von Instanzen eines anderen Attributs. Dies ist bei den Tabellen `t_definitions`, `t_varianten`, `t_äquivalent`, `t_beispiele` und `t_verweise` der Fall: Z.B. kann ein Terminus die Variante mehrerer Lemmata sein, so wie ein Lemma mehrere Varianten haben kann.

An dieser Stelle soll noch einmal auf die Anwendung und Nützlichkeit der Normalisierung von Tabellen ausgehend von einer Gesamttabelle eingegangen werden. Bei einer Tabelle in der ersten Normalform sind die Wertebereiche aller Attribute atomar. Das bedeutet, dass jedes Feld eines jeden Attributs einen einzigen Wert enthalten darf: jede Definition, jedes Beispiel, jede Variante usw. muss in ein separates Feld eingetragen werden. Das hat zur Folge, dass jeder einzelne Wert unabhängig in der Datenbank gesucht werden kann.

Eine Tabelle ist in der zweiten Normalform, wenn sie sich in der ersten Normalform befindet und jedes Nichtschlüsselattribut voll funktional abhängig von jedem Schlüsselkandidaten ist, d.h. es ist vom gesamten Schlüsselkandidaten funktional abhängig, aber nicht bereits von Teilen davon. Eine funktionale Abhängigkeit liegt wiederum vor, wenn zu jedem Wert eines Attributs höchstens ein Wert von einem anderen Attribut vorhanden ist. Bei unserem Experiment ist die zweite Normalform bereits gegeben, deshalb wurde direkt zur dritten Normalform übergegangen.

Die dritte Normalform besagt, dass eine Tabelle in der zweiten Normalform sein muss und keine transitiven Abhängigkeiten zwischen Nichtschlüsselattributen und Schlüsselkandidaten bestehen dürfen. Das führt zur Aufteilung von Tabellen, wenn ein Attribut (z.B. Quelle) nicht unmittelbar vom Schlüsselkandidaten (z.B. TermID) abhängig ist, sondern von einem anderen Attribut (z.B. Definition). Diese Regel stellt sicher, dass die Tabellen einer relationalen Datenbank nur direkte Abhängigkeiten enthalten.

In der Praxis finden am häufigsten die ersten drei Normalformen Anwendung, da sie die wesentlichen Problematiken ansprechen. Während des Normalisierungsprozesses besteht die Möglichkeit, weitere Regeln anzuwenden, ein Thema, auf das in diesem Beitrag jedoch nicht eingegangen wird. Wie in Abschnitt 2.4 erwähnt, findet die Umwandlung der Tabellen in die ersten drei Normalformen während der Erstellung des Datenmodells statt.

4.2. Erstellung von Einträgen

Nachdem die Anzahl und Struktur der einzelnen Tabellen festgelegt und die Zusammenhänge zwischen den verschiedenen Tabellen in der Datenbank anhand von Schlüsseln sichergestellt wurden, werden alle Tabellen mit den zuvor gesammelten und aufbereiteten lexikalischen und konzeptuellen Daten gefüllt. Die Tabellen werden im CSV-Format angelegt, einem Tabellenformat, das sich zum Import oder zur direkten Erstellung im Datenbankmanagementsystem eignet.

Zu jeder Tabelle werden nun Auszüge aus den Daten für die Fachsprache der Immunologie angezeigt. Aus Gründen der Lesbarkeit wird die Tabelle `t_haupttabelle` im Folgenden aufgeteilt.

(1) t_haupttabelle:

TermID	Terminus	Sprache	IPA	Wortklasse	Genus	Genitiv	Plural
1	anticorpi	IT		N	m.		(-)
2	agente pato- geno	IT		N + A	m.		agenti pato- geni
3	Antikörper	DE		N	m.	Antikörper	(-)
4	impfen	DE		V			
5	COVID-19	DE	[,ko:vɪt'noɪn tse:n]	N	m./n./f.	COVID-19	
6	COVID-19	IT		N	m.		

TermID	Kommunikative Ebene	Konzept	Abbildung
1	esperti/non-esperti	SISTEMA IMMUNITARIO	
2	esperti	AGENTI PATOGENI	https://st3.depositphotos.com/2731675/35262/i/600/depositphotos_352621372-stock-photo-conceptual-illustrative-virus-image-of.jpg
3	Experten/Nicht-Experten	IMMUNSYSTEM	https://static.spektrum.de/fm/912/thumbnails/6813384933_166d8e7f29_o.jpg.3113806.jpg
4	Experten/Nicht-Experten	PROZESSE	
5	Experten/Nicht-Experten	KRANKHEITEN	
6	Experten/Nicht-Experten	PATOLOGIE	

Felder, die sich auf ein bestimmtes Attribut beziehen, können den Null-Wert annehmen bzw. leer bleiben, wenn kein Wert anzugeben ist (z.B. die Genitivform bei italienischen Substantiven). Im Falle von Lemmata, die hauptsächlich im Plural verwendet werden, z.B. *Antikörper*, wird beim Attribut „Plural“ die Angabe „(-)“ gemacht. Die Angaben werden jeweils in der Sprache des Lemmas gemacht, auch im Falle der Konzepte. Für unser Experiment wurden sechs Konzepte ausgewählt, die den gesamten Fachwortschatz der ausgewählten Texte abdecken: CORPO E ORGANI / KÖRPER UND ORGANE, SISTEMA IMMUNITARIO / IMMUNSYSTEM, PROCESSI / PROZESSE, PATOLOGIA / KRANKHEITEN, AGENTI PATOGENI / KRANKHEITSERREGER, AGENTI / AGENTEN.

(2) t_definitionen:

TermID	Definition
1	Proteina diretta contro una sostanza riconosciuta come estranea (antigene) che contribuisce alla difesa immunitaria dell'organismo.
2	Microrganismo in grado di causare malattie.
3	Antikörper oder Immunglobuline sind globuläre <u>Proteine</u> , die von zu Plasmazellen differenzierten B-Lymphozyten produziert und sezerniert werden. Sie sind gegen Bestandteile eines Antigens gerichtet und besitzen die Fähigkeit an dieses zu binden.

Einem Lemma entspricht meistens eine einzige Definition. In manchen Fällen ist es allerdings notwendig, Teildefinitionen bzw. Definitionen aus unterschiedlichen Quellen anzugeben.

(3) t_definitionsquellen:

Definition	Quelle
Proteina diretta contro una sostanza riconosciuta come estranea (antigene) che contribuisce alla difesa immunitaria dell'organismo.	http://www.paginesanitarie.com/anticorpo.htm
Microrganismo in grado di causare malattie.	http://www.paginesanitarie.com/agentepatogeno.htm

Antikörper oder Immunglobuline sind globuläre <u>Proteine</u> , die von zu Plasmazellen differenzierten B-Lymphozyten produziert und sezerniert werden. Sie sind gegen Bestandteile eines Antigens gerichtet und besitzen die Fähigkeit an dieses zu binden.	https://flexikon.doccheck.com/de/Antik%C3%B6rper
--	---

In einem späteren Schritt ist es möglich, Definitionen mit Identifikationsnummern zu versehen. Dies würde allerdings die Erstellung einer neuen Tabelle mit den Attributen „Definition“ und „DefinitionID“ erfordern. Darauf wurde hier aus Gründen besserer Nachvollziehbarkeit verzichtet.

(4) t_varianten:

TermID	Variante	Variationstyp
2	agente infettivo	MV/ SV- OV-
3	Immunglobuline	MS+ SV- OV-
6	Covid-19	MV- SV- OV+

Wie in Abschnitt 3 erläutert, beschreibt das Attribut „Variationstyp“ die Kombination aus einem bestimmten Terminus und einer spezifischen Variante. Um die Konstellation der drei Variationsmöglichkeiten (morphologisch, syntaktisch und orthografisch) anzugeben, verwenden wir die oben vorgeschlagene Notation.

(5) t_äquivalente:

TermID	Äquivalent
1	Antikörper
2	Krankheitserreger
2	Erreger
3	anticorpo
4	vaccinare
5	COVID-19
5	Covid-19

Ein Lemma darf mehrere Äquivalente haben – z.B. hat der Terminus *agente patogeno Krankheitserreger* und *Erreger* als Äquivalente. Es ist hier kein Null-Wert zulässig: Falls ein Lemma keine Entsprechung in der Zielsprache hat, soll beispielsweise eine Paraphrase angeboten bzw. der Terminus aus der Ausgangssprache als Fremdwort übernommen werden.

(6) t_beispiele:

TermID	Beispiel
1	L'AIFA ha aggiornato le modalità di utilizzo degli anticorpi monoclonali anti COVID-19 in relazione alle nuove evidenze di letteratura che si sono rese recentemente disponibili.
2	La percentuale di popolazione che deve divenire immune per proteggere anche chi è ancora suscettibile varia a seconda dell'agente patogeno.
4	Wie sollen Personen sich impfen lassen, die an klinischen Impfstoffstudien teilgenommen haben?

Ähnlich wie bei Definitionen dürfen einem Lemma mehrere Beispiele zugeordnet werden.

(7) t_beispielquellen:

Beispiel	Quelle
L'AIFA ha aggiornato le modalità di utilizzo degli anticorpi monoclonali anti COVID-19 in relazione alle	https://www.aifa.gov.it/-/definizione-delle-modalit%C3%A0-ottimali-d-uso-degli-anticorpi-monoclonali-anti-covid-19

nuove evidenze di letteratura che si sono rese recentemente disponibili.	
La percentuale di popolazione che deve divenire immune per proteggere anche chi è ancora suscettibile varia a seconda dell'agente patogeno.	https://www.iss.it/documents/20126/0/Glossario.pdf/fe8c209d-33b7-dbc4-b324-44eee879895f?t=1585325504835
Wie sollen Personen sich impfen lassen, die an klinischen Impfstoffstudien teilgenommen haben?	https://www.rki.de/SharedDocs/FAQ/COVID-Impfen/FAQ_Liste_Durchfuehrung_Impfung.html

Auch hier können Identifikationsnummern für die Beispiele (BeispielID) vergeben werden.

(8) t_verweise:

TermID	Verweis
2	6
5	4
6	2

Verweise erfolgen zwischen Lemmata der gleichen Sprache und werden daher in der Tabelle t_verweise nur durch die TermID dieser Lemmata identifiziert. In den vorliegenden Daten wird vom italienischen Lemma *agente patogeno* auf das italienische Lemma *COVID-19* verwiesen. Je nach Fall erfolgt die Verweisung unidirektional (nur eines der Lemmata fungiert als Verweisziel – hier *impfen*) oder bidirektional (beide Lemmata sind Verweisziele – hier *agente patogeno* und *COVID-19*).

Ein terminologischer Eintrag besteht aus der Vereinigung der Datensätze, die sich auf ein bestimmtes Lemma beziehen und auf die verschiedenen Tabellen verteilt sind. Zur Veranschaulichung sind in allen Tabellen die Datensätze, die sich auf den Terminus (Lemma) mit TermID 2 beziehen, grau hinterlegt. Weitere Attribute wären notwendig, um die lexikografischen Einträge zu vervollständigen, insbesondere wenn bestimmte Lemmata mehr Details erfordern. Nützlich wären z.B. grammatikalische Angaben und Disambiguateure zu den Äquivalenten, die Kennzeichnung von Wortfamilien (*impfen, Impfung, Impfstoff, ...*), pragmatische Angaben verschiedener Art, Anmerkungen bei unregelmäßigem Gebrauch und Phänomenen wie Pluralia tantum (vgl. *Antikörper*) sowie die Angabe des Bearbeitungsstatus eines Eintrags.

Die Datenbank wird im phpMyAdmin-Datenbankmanagementsystem implementiert, das auf der Open-Source-Plattform Xampp, Version 8.0.10 (<https://www.apachefriends.org>) verfügbar ist. Nachdem eine Datenbank erstellt und benannt wurde, können die Tabellen direkt in phpMyAdmin erstellt und strukturiert und im Anschluss mit Daten gefüllt werden, oder sie können im CSV-Format erstellt, ausgefüllt und in phpMyAdmin importiert werden.

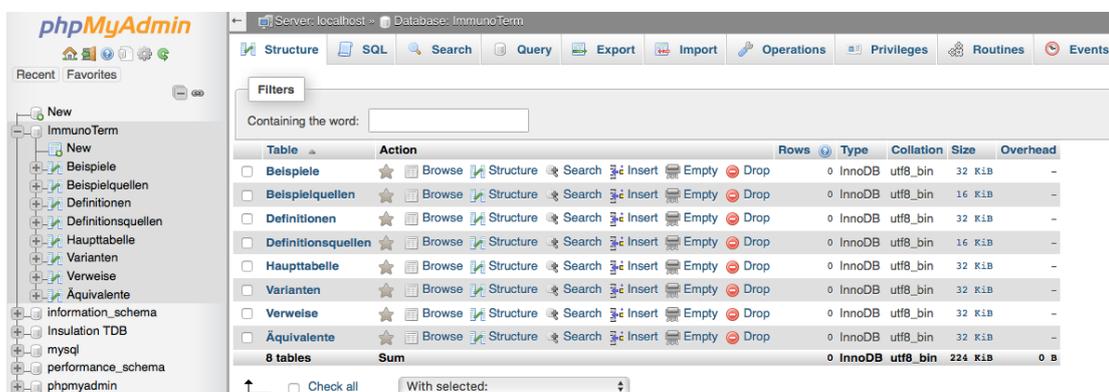


Abbildung 3
Struktur von ImmunoTerm im phpMyAdmin-Datenbankmanagementsystem.

Abbildung 3 ist ein Screenshot von ImmunoTerm in phpMyAdmin, der die allgemeine Struktur der Datenbank für immunologische Sprachen zeigt. Tabellen und Collations sind sichtbar. Collations sind Regeln in der Datenbank für die Sortierung und den Vergleich von Daten im SQL-Server. Abbildung 4 zeigt die Struktur der Tabelle t_haupttabelle, aus der man die den einzelnen Attributen zugeordneten Datentypen, den Primärschlüssel und weitere Indizes erkennen kann.

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
<input type="checkbox"/>	1 TermID	int(11)			No	None		AUTO_INCREMENT	Change Drop Primary Unique More
<input type="checkbox"/>	2 Terminus	varchar(128)	utf8_bin		No	None			Change Drop Primary Unique More
<input type="checkbox"/>	3 Sprache	varchar(32)	utf8_bin		No	None			Change Drop Primary Unique More
<input type="checkbox"/>	4 IPA	varchar(64)	utf8_bin		No	None			Change Drop Primary Unique More
<input type="checkbox"/>	5 Wortklasse	varchar(32)	utf8_bin		No	None			Change Drop Primary Unique More
<input type="checkbox"/>	6 Genus	varchar(32)	utf8_bin		No	None			Change Drop Primary Unique More
<input type="checkbox"/>	7 Genitiv	varchar(128)	utf8_bin		Yes	NULL			Change Drop Primary Unique More
<input type="checkbox"/>	8 Plural	varchar(128)	utf8_bin		Yes	NULL			Change Drop Primary Unique More
<input type="checkbox"/>	9 Kommunikative Ebene	varchar(128)	utf8_bin		No	None			Change Drop Primary Unique More
<input type="checkbox"/>	10 Konzept	varchar(128)	utf8_bin		No	None			Change Drop Primary Unique More
<input type="checkbox"/>	11 Abbildung	varchar(256)	utf8_bin		Yes	NULL			Change Drop Primary Unique More

Check all With selected: Browse Change Drop Primary Unique Index Add to central columns Remove from central columns

Abbildung 4

Struktur der Haupttabelle in ImmunoTerm. Es sind u.a. Attribute, Datentypen von Attributen und Schlüssel sichtbar.

Aus den letzten Abbildungen geht hervor, dass Deutsch als Metasprache verwendet wurde, z.B. bei der Benennung der Tabellen und der Attribute der einzelnen Tabellen. Diese Wahl ist auf Überlegungen zurückzuführen, die mit dem Kurs zusammenhängen. Die Metasprache hat allerdings keinen Einfluss auf den Inhalt und die Struktur der Terminologiedatenbank, die die gleiche Art der Arbeit durch Muttersprachler und Nicht-Muttersprachler voraussetzt.

Es gibt inzwischen verschiedene Methoden zur Erstellung einer lexikografischen Ressource auf der Grundlage der relationalen Terminologiedatenbank. Die aufwendigste und zeitintensivste Methode, gleichzeitig aber auch die geeignetste für fachlexikografische Projekte, mit denen auch unter dem Gesichtspunkt des Wörterbuchdesigns eine neue Ressource geschaffen werden soll, ist die Erstellung einer Weboberfläche, die die Daten der Termbank über php/SQL einliest und individualisiert bzw. dynamisch darstellt. Darüber hinaus können phpMyAdmin-Tabellen in verschiedenen Formaten exportiert werden, einschließlich XML. Dies bedeutet, dass es möglich ist, Terminologiedaten in ein Wörterbuchredaktionssystem zu importieren, das XML unterstützt.

5. Fazit: Die Arbeit mit fachsprachlichen Korpora und ihr Mehrwert für den Übersetzungsunterricht

Bei dem hier beschriebenen Experiment liegt der theoretische Schwerpunkt auf dem Thema der terminologischen Variation. Die Methode ist jedoch hinreichend flexibel, um die Untersuchung anderer Phänomene zu ermöglichen, z.B. fachsprachlicher Kollokationen, lexikografischer und terminografischer Definitionen sowie semantischer Relationen. Darüber hinaus besteht die Möglichkeit, die Anzahl der in der Datenbank erfassten Sprachen zu erweitern, indem beispielsweise Englisch als einzige Ausgangssprache neben Italienisch und Deutsch als Zielsprache aufgenommen wird. Bei mehrsprachigen Datenbanken kann Englisch ggf. als Metasprache eingesetzt werden, etwa zur Bezeichnung von Klassen- und Attributnamen.

Wie bereits erwähnt, ist das Experiment für Muttersprachler/Muttersprachlerinnen und Nicht-Muttersprachler/Nicht-Muttersprachlerinnen gleichermaßen geeignet. Im Falle einer zweisprachigen relationalen Datenbank wird diese in der Regel in beide Sprachrichtungen entwickelt. Somit ist die

Datenbank bidirektional. Gleichzeitig ist sie bifunktional, d.h. sie unterstützt sowohl das passive als auch das aktive Übersetzen. Dies bedeutet, dass an beiden Korpora die gleiche Art von Analyse durchgeführt wird.

Eine Lösung, die sich als besonders interessant erwiesen hat, ist die Kombination der Übung zur Erstellung von Terminologiedatenbanken mit einem Fachübersetzungskurs unter Verwendung der entsprechenden Ausgangs- und Zieltexte. Auf diese Weise kann die Nützlichkeit solcher auf die Bedürfnisse von Übersetzern zugeschnittenen Datenbanken beim Übersetzungsprozess direkt demonstriert werden.

Die Erstellung einer relationalen Datenbank in einer Umgebung wie Xampp ist der erste Schritt zur Schaffung einer lexikalischen Online-Ressource. Obwohl dies nicht im Rahmen eines einzigen Kurses erreicht werden kann, erhalten die Studierenden Informationen über den gesamten lexikografischen bzw. terminografischen Prozess. In diesem Zusammenhang soll noch einmal auf die Nützlichkeit von *Sketch Engine* im Unterricht hingewiesen werden: Die benutzerfreundliche Online-Plattform mit ihren unterschiedlichen Korpusabfragetools ermöglicht die Anwendung einfacher Pipelines bei der Durchführung von korpusbasierten Experimenten. Darüber hinaus bieten Redaktionssysteme für Lexikografie und Terminografie, etwa *Lexonomy* und *Terminologue*, zusätzliche Möglichkeiten der Erstellung von fertigen Ressourcen basierend auf einer terminologischen Datenbank.

Die Arbeit mit fachsprachlichen Korpora ist nicht nur ein unverzichtbares Werkzeug bei der Erstellung von Ressourcen zur Unterstützung der Übersetzung. Sie kann in den Übersetzungsunterricht integriert werden, um die Prozesse zu veranschaulichen, die mit der Erstellung solcher Ressourcen verbunden sind. Bei der Umsetzung dieser Prozesse ermöglicht die Korpusarbeit angehenden ÜbersetzerInnen, theoretisch relevante Aspekte der Fachkommunikation wie das Phänomen der terminologischen Variation, anhand von Textdaten in ihrer Mutter- oder Fremdsprache zu erforschen.

6. Literatur und Ressourcen

Arntz, Reiner / Picht, Herbert / Schmitz, Klaus-Dirk (2014): *Einführung in die Terminologearbeit*. Hildesheim: Georg Olms Verlag.

Bergenholtz, Henning / Tarp, Sven / Wiegand, Herbert Ernst (2008): Daten Distributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. In: Hoffmann, Lothar et al. (Hrsg.): *Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft*, 2. Halbband. Berlin, New York: De Gruyter Mouton, 1762-1832.

Busch, Albert / Spranz-Fogasy, Thomas (Hrsg.) (2015): *Handbuch Sprache in der Medizin*. Berlin: De Gruyter Mouton.

Cortelazzo, Michele A. (2009): *Lingue speciali. La dimensione verticale*. Padova: Unipress.

Faber, Pamela (2015): Frames as a framework for terminology. In: *Handbook of terminology* 1: 14, 14-33.

Giacomini, Laura (2018): Frame-based Lexicography: Presenting Multiword Terms in a Technical E-dictionary. In: *Proceedings of the XVIII EURALEX International Congress. Lexicography in Global Contexts*, 309-318.

Giacomini, Laura (2019). *Ontology – Frame – Terminology. A method for extracting and modelling variants of technical terms*. Habilitationsschrift, Universität Hildesheim.

Giacomini, Laura (2021, i.E.): Terminologearbeit und Wissensmanagement. In: Atayan, Varham / Metten, Thomas / Schmidt, Vasco Alexander (Hrsg.): *Handbuch Sprache in Mathematik, Naturwissenschaften und Technik. Handbücher Sprachwissen (HSW)*, Band 15. De Gruyter.

Kilgarriff, Adam et al. (2014): The Sketch Engine: Ten years on. In: *Lexicography* 1, 7-36.

Pera, Franz / Schmiedebach, Hans-Peter (2007): *Medizinischer Wortschatz: Terminologie kompakt*. Berlin: de Gruyter.

Serianni, Luca (2005): *Un treno di sintomi. I medici e le parole: percorsi linguistici nel passato e nel presente*, Garzanti.

Biographische Notiz: Laura Giacomini lehrt und forscht am Institut für Übersetzen und Dolmetschen (IÜD) der Universität Heidelberg und am Institut für Informationswissenschaft und Sprachtechnologie (IWiST) der Universität Hildesheim. Sie hat einen PhD in Angewandter Linguistik und eine Habilitation in Fachkommunikation und NLP. Ihre Forschungsinteressen umfassen die Bereiche (computergestützte) Terminologie und Lexikographie, Wissensrepräsentation, Phraseologie sowie korpuslinguistische Ansätze zur Translationswissenschaft.

Kontaktanschrift:

Dr. Laura Giacomini
Universität Heidelberg
Institut für Übersetzen und Dolmetschen
Plöck 57a
69117 Heidelberg
laura.giacomini@iued.uni-heidelberg.de

