

ALICJA SZTUK
(UNIwersytet Warszawski, Warszawa)

TERMINOLOGIEDATENBANKEN – MÖGLICHKEITEN DER QUALITÄTSAUSWERTUNG

ABSTRACT

Terminology is significant for professional communication and *ipso facto* for translation quality assurance (QA). To deliver a translation of high quality, it is crucial to have all new terms that occur in professional discourse collected, stored and managed properly by means of terminology databases (TDBs). In this paper I will try to define ‘quality’ in relation to TDBs and to determine the methodology and criteria that need to be considered by evaluating a TDB in the context of its reliability.

KEYWORDS: terminology, terminology database (TDB), translation process, terminology management, quality assurance (QA)

STRESZCZENIE

Terminologia jest kluczowa w komunikacji specjalistycznej, a tym samym również w zapewnieniu jakości tłumaczenia. Aby dostarczyć tłumaczenie najwyższej jakości istotne jest, aby terminy pojawiające się w dyskursie specjalistycznym były zbierane, przechowywane i zarządzane przy pomocy baz danych terminologicznych. W artykule podejmę próbę zdefiniowania „jakości” w odniesieniu do baz danych terminologicznych oraz próbę określenia metodologii i kryteriów oceny danej bazy terminologicznej w kontekście jej jakości (wiarygodności).

SŁOWA KLUCZOWE: terminologia, baza danych terminologicznych, proces tłumaczenia, zarządzanie terminologią, zapewnienie jakości

EINLEITUNG

Der vorliegende Beitrag widmet sich im Allgemeinen den folgenden Themen: Terminologie, Terminologieverwaltung¹ und insbesondere Terminologiedatenbanken (auch kurz Termbanken genannt). Es wird davon ausgegangen, dass die effektive Terminologieverwaltung einen wichtigen Beitrag zum konsistenten Übersetzen leistet.

¹ In der Fachliteratur und in der Übersetzungsbranche wird Terminologieverwaltung synonymisch auch als Terminologiemanagement bezeichnet. Dabei scheint sich der zweite Ausdruck größerer Popularität (besonders im Bereich der Übersetzungsdienstleistungen) zu erfreuen.

Diesbezüglich steht dieses Problem (Terminologieverwaltung) im Vordergrund meiner Erörterungen. Dieser erwähnte Ausdruck bezieht sich einerseits auf die Qualität der ganzen terminologischen Infrastruktur, andererseits hat er einen wesentlichen Einfluss auf die Qualitätssicherung (QS²) im Übersetzungsprozess. Die effektive Terminologieverwaltung soll in erster Linie durch zuverlässige (von höchster Qualität) Termbanken sichergestellt werden. Meine Erwägungen über qualitative Aspekte der terminologischen Verwaltungsprozesse werden durch die Erläuterung des Basisausdrucks Terminologie eingeleitet. Diese Einleitung ist für alle nachfolgenden Überlegungen grundlegend, da der Ausdruck Terminologie nicht eindeutig ist, sich in verschiedenen Kontexten unterschiedlich gebrauchen lässt, und durch Begriffsdiskrepanzen zu unerwünschten Missverständnissen führen kann. Im Weiteren wird auf die Rolle der Terminologie im fachlichen Diskurs und dadurch auch auf ihre Rolle in der Qualitätssicherung beim Übersetzungsprozess hingewiesen. Anschließend wird zum Leitthema dieses Beitrags – Terminologiedatenbanken und ihre Qualität – übergegangen.

Der laufende Zuwachs neuer Termini in jedem Bereich menschlicher Tätigkeit verlangt die Erstellung und zugleich Aktualisierung der Termbanken, die meines Erachtens der wichtigste Bestandteil der terminologischen Infrastruktur darstellen. Immer neu aufkommende Termini müssen fortlaufend gesammelt, geordnet, mit entsprechenden Metadaten versehen und schließlich effektiv verwaltet werden, sonst wird der gewünschte Verlauf des Kommunikations- und des Übersetzungsprozesses gefährdet. Das Ziel jeglicher Terminologieverwaltung besteht darin, sicherzustellen, dass der Schlüsselwortschatz für ein Projekt, eine Organisation, ein Produkt usw., in einem für alle Projektpartner zugänglichen System aufbewahrt wird, das regelmäßig aktualisiert wird und für alle am Prozess Beteiligten jederzeit leicht zugänglich ist und von allen, abgesehen von der Projektphase, einheitlich verwendet wird.

Deswegen wird der Schwerpunkt des vorliegenden Beitrags auf die Qualität (Zuverlässigkeit) der terminologischen Infrastruktur, insbesondere der Terminologiedatenbanken, gelegt. Es wird versucht, die universelle Methodologie und Kriterien auszuarbeiten, die die Qualitätsauswertung jeder Terminologiedatenbank ermöglichen. Anschließend wird eine vergleichende (kontrastive) Analyse von zwei Terminologiedatenbanken (Terminologieverwaltungssystemen – TVS³) vorgenommen, bei der nicht nur auf den Inhalt (einbezogene Rekorde – Termini) einer Termbank eingegangen wird, sondern auch auf andere Komponenten, die für die Qualität und dadurch für die Zuverlässigkeit einer Bank entscheidend sind. Zu

² Qualitätssicherung (QS) – eng. Quality Assurance (QA), oder Qualitätskontrolle – eng. Quality Control (QC) ist ein Sammelbegriff für unterschiedliche Ansätze und Maßnahmen zur Sicherstellung festgelegter Qualitätsanforderungen (<https://de.wikipedia.org/wiki/Qualit%C3%A4tssicherung>). Im Falle der computerunterstützten Übersetzung kann QS auch als ein mehrstufiger Prozess bezeichnet werden, der in einem übersetzten Dokument nach maschinell erkennbaren Fehlern sucht und dadurch eine bestimmte Qualität sicherstellt.

³ TVS sind die heutzutage eingerichteten Softwares, die die vorhandenen Terminologieressourcen komplex pflegen, verwalten und sie, wenn erforderlich, bereitstellen.

diesen gehören u.a. der Aufbau, die Struktur einer Termbank, die Vorgehensweise bei dem Aktualisierungsprozess sowie die Methoden der Verarbeitung, Pflege und Bereitstellung von Terminologie. Die erwähnte Analyse soll veranschaulichen, inwiefern die effektive Terminologieverwaltung für die Qualitätssicherung im Übersetzungsprozess unerlässlich ist und schließlich, inwiefern sie tatsächlich die Zuverlässigkeit der Terminologie gewährleistet.

Abschließend wird der Versuch unternommen, Qualität *stricte* in Bezug auf Terminologiedatenbanken zu definieren, indem die Merkmale genannt werden, die ihre Qualität kennzeichnen. Infolgedessen lassen sich die konkreten Anforderungen an Termbanken der Zukunft (deren Erstellung, sowie deren Aktualisierung) bestimmen.

TERMINOLOGIE – EINGRENZUNG DES ENTITÄTEN- UND FORSCHUNGSBEREICHES

An dieser Stelle muss explizit darauf hingewiesen werden, dass Terminologie nicht das Gleiche wie eine Fachsprache (Fachsprache eines bestimmten Fachbereichs) ist und, dass keine Fachsprache der Terminologie gleichgesetzt werden kann. Infolgedessen dürfen diese Ausdrücke nicht synonymisch verwendet werden. Zwar stellt die Terminologie einen wesentlichen Bestandteil jeder einzelnen Fachsprache dar, sie ist jedoch nicht die einzige Komponente davon. Die Fachlichkeit einer Sprache kennzeichnet sich auch durch andere Merkmale, sowohl auf der Satz-, als auch auf der Textebene. In diesem Sinne darf die Terminologie nicht den Fachsprachen gleichgesetzt werden. Erstere ist lediglich im engeren Sinne als Fachsprache aufzufassen. Wie soll also Terminologie theoretisch betrachtet, definiert und erörtert werden? Bevor ein konkreter Definitionsversuch angegangen wird, muss in erster Linie betont werden, dass Terminologie zwei sich wesentlich voneinander unterscheidende Phänomene repräsentieren kann, d. h. dem Ausdruck *Terminologie* können zwei Bedeutungen zugeschrieben werden. Erstens kann Terminologie für die Bezeichnung einer Sammlung von terminologischen Einheiten (Termini) gebraucht werden. Andererseits kann Terminologie auch als Wissensgebiet bezeichnet werden, im Rahmen dessen man sich wissenschaftlich mit Termini beschäftigt: mit deren Bildung, Verwaltung, Harmonisierung und Normung.

Sowohl in diesem als auch im vergangenen Jahrhundert sind zahlreiche Beiträge entstanden, die zum Ziel hatten, das Phänomen Terminologie zu erforschen, zu erklären, zu erörtern und schließlich zu definieren und zu beschreiben. Mit der Zeit wurden in diesem Bereich verschiedene wissenschaftliche Ansätze⁴ entwickelt.

⁴ Wüster (1979), Gajda (1990), F. Gucza (1991, 2002), S. Gucza (2002, 2004, 2008, 2010), Jurkowski (1991), Lukszyn (1991, 1998 und 2001 – mit Zmarzer), Kockaert/ Steurs (2015), Nowicki (1986), Wright mit Budin (1997), Sztuk (2017) und viele andere.

Sie haben zweifelsohne feste Grundsteine für die heutige Terminologielehre gelegt, obwohl auch viele von den terminologischen Ansätzen und Theorien heute als veraltet gelten und sogar als Gegenbeispiele in der Terminologielehre vorgestellt werden.

TERMINOLOGIE IM FACHDISKURS UND IHRE ROLLE IN DER QUALITÄTSSICHERUNG BEIM ÜBERSETZEN

Die erwähnten Wissenschaftler haben mehr oder weniger diverse terminologische Ansätze entwickelt und repräsentiert. Sie waren jedoch miteinander einig bezüglich der bedeutenden Rolle, die die Termini in der Fachkommunikation (im Fachdiskurs) spielen: Terminologie beeinflusst und prägt weitgehend die Fachkommunikation. Ohne Terminologie ist keine Fachsprache aufzufassen.

F. Grucza hat es mehrmals in seinen Beiträgen betont, dass Fachtexte und zugleich auch die dort verwendeten Termini, einen determinierenden Faktor der Zivilisationsentwicklung darstellen (vgl. F. Grucza 2008: 18). Immer mehr neue Termini, die gebildet und verwendet werden, führen dazu, dass der Mensch nicht mehr imstande ist, die ganze Terminologie sogar aus einem bestimmten Fachbereich weder aktiv noch passiv zu operationalisieren. Infolgedessen kann der Mensch diese Terminologie (ohne z. B. Softwareunterstützung) nicht unter Kontrolle haben und effektiv verwalten. Die irrtümlicherweise gepflegte und falsch verwaltete Terminologie, kann daher nicht zur Qualitätssicherung beim Übersetzen beitragen. Eine logische Schlussfolgerung daraus lautet: die Qualitätssicherung im Übersetzungsprozess ist wesentlich auf die Qualität der terminologischen Infrastruktur zurückzuführen.

Qualitätssicherung wird in der Fachliteratur wie folgt definiert: „QA (Quality Assurance) is a model approach that ensures good results if the right combination of human and technical resources is used in a sequence of steps and tasks that constitute a process within a system” (Popiotek 2015). Diese Definition lässt sich selbstverständlich auch auf den Übersetzungsprozess beziehen, indem bestimmte Aufgaben und Maßnahmen in einer festgelegten Reihenfolge vorgenommen werden, um die gewünschte Übersetzungsqualität sicherzustellen. Eine dieser Aufgaben ist die kohärente, einheitliche und präzise Terminologie in der Übersetzung anzuwenden.

Daraus lässt sich schlussfolgern, dass die Qualität eines (Fach)Textes von der Qualität der verwendeten Terminologie abhängig ist. Immer mehr Wirtschaftssubjekte, die sich gewissermaßen auf die Übersetzungsbranche (TSPs – *Translation Service Providers*) verlassen, wünschen sich, dass die Terminologie für ihre Projekte, Marke usw. entsprechend gesammelt, gepflegt und verwaltet wird, damit alle Texte kohärent übersetzt werden können und im Ergebnis die gewünschte Qualität

haben. Aufgrund langjähriger Zusammenarbeit entstehen üblicherweise nicht nur Termbanken, sondern auch TMs – *Translation Memories*, die sowohl mehrsprachige Terminologie speichern, als auch ganze Textsegmente (Sätze oder vollständige Textabschnitte) in der Ausgangs-, sowie in der Zielsprache. *Translation Memories* arbeiten in einem CAT-Tool im Hintergrund und unterbreiten die Vorschläge für die Übersetzung einzelner Segmente, die identisch oder ähnlich sind. Solche Vorschläge können nur dann zur Verfügung gestellt werden, wenn Terminologie angemessen angepasst wird.

Einheitliche Terminologie wird vor allem bei großen Übersetzerteams zur Herausforderung, da die Projektmanagers (PMs) sicherstellen müssen, dass alle am Projekt beteiligten Übersetzer dieselbe Terminologie verwenden. Die PMs sind außerdem dafür zuständig, dass die Texte in der Zielsprache konsistent zueinander sind und dass sie den Kunden in der höchsten Qualität noch vor dem Ablauf der Deadline geliefert werden.

Die erwähnten Maßnahmen stellen nur einen geringen Teil von Prinzipien dar, die in der Terminologieverwaltung gelten. Es sind mehrere Richtlinien und (ISO)-Normen entstanden, die die beste Praxis und Modelle im Bereich Terminologieverwaltung aus verschiedenen Perspektiven abdecken. Dazu gehört u. a. die ISO Norm 704 *Terminology. Principles and Methods*, die allgemeine Richtlinien dafür angibt, wie man mit Terminologie vorgehen soll. Es gibt auch die ISO-Normen 12616 und 12620, entsprechend über translationsorientierte Terminographie und Spezifikation von Datenkategorien für terminologische Einträge. In Hinsicht auf die sogenannte Interoperationalität und hinsichtlich verschiedener Verwaltungstools, Diskrepanzen in Datenkategorien und Speicherformaten der Einträge wurde die ISO-Norm 30042 *TermBase eXchange (TBX)* bearbeitet. Dieses Dokument kommt der Vielfalt der Speicherformate entgegen und empfiehlt, wie man vorgehen soll, wenn man terminologische Daten zwischen den Termbanken austauschen will und sich keinen Datenverlust oder deren Attribute wünscht.

Um auf die Rolle der Terminologie in der Qualitätssicherung konkret hinzuweisen, sollen in erster Linie die einzelnen Standardetappen (Bestandselemente) des vollständigen Prozesses der Qualitätssicherung beim Übersetzen genannt werden. Diese Etappen möchte ich nach Popiołek (2015) anführen:

1. Defined supplier (translator, reviser, reviewer) recruitment process;
2. Procedures for handling quotations, client documents, contracts, project specifications;
3. Set competences and experience required of suppliers (translators, revisers, reviewers) and procedures for job assignment;
4. Well defined and documented procedures for managing terminology and other technical and linguistic resources (TM, MT, style-guides, lexicons, reference materials, etc.);
5. Well defined and documented procedures for managing all stages of the translation process itself (ideally through some sort of automated work-flow);
6. Ongoing quality control system that always involves professional revision of all the translated material (source and target content comparison segment by segment and

assessment according to reliable metrics), additional review if required and regular feedback;

7. Procedures for additional random spot checks to ensure the control system ensures maintenance of required quality levels and additional assurance measures;
8. Procedures for dealing with complaints, reporting and implementing corrective measures;
9. Procedures for ensuring data security and archiving.

Die Autorin betont die Tatsache, dass die durch ein CAT-Tool erkannten Fehler kategorisiert werden können. Auf diese Art und Weise ist es u. a. möglich auch terminologische Fehler zu finden. Zu dieser Fehlergruppe gehören z. B. Termini, die in einem falschen Kontext gebraucht wurden, oder wenn z. B. für einen Gegenstand mehrere Ausdrücke synonymisch verwendet werden. Diese Tatsache bestätigt die Bedeutung der Terminologie und deren Kohärenz bei dem komplexen Übersetzungsprozess sowie bei dem auf ihn bezogenen Qualitätssicherungsprozess.

Die erwähnten Etappen geben nur ganz allgemein die nacheinander folgenden Schritte der Vorgehensweise an. Dennoch kann man die wichtige Rolle der einheitlichen Terminologie auf den ersten Blick erkennen, insbesondere in den Punkten 4–6, wo die Terminologie entweder direkt genannt wird oder indirekt in dem Ausdruck *translation* oder *translation process* enthalten ist. Wichtig dabei ist auch, dass der Übersetzungsprozess von zwei unabhängigen Prüfern kontrolliert und, falls erforderlich, korrigiert wird. Diesbezüglich wird auch die verwendete Terminologie zweimal geprüft und eventuell nachgebessert. Alles soll in einem CAT-Tool ablaufen, so dass sogar kleine Tippfehler vermieden werden können. Der Einsatz eines CAT-Tools soll übrigens umgehend dem Übersetzer (auch den beiden Prüfern) die einheitliche, vorbestimmte und vor allem genehmigte Terminologie vorschlagen. Diese Übersetzungserleichterung besteht darin, dass diese vorgeschlagenen Rekorde vorher von jemandem in einer Termbank oft in einem bestimmten Kontext gespeichert wurden und jetzt von dem CAT-Tool als Termvorschlag erkannt werden. Diese Vorschläge erleichtern den Übersetzungsprozess und sichern, dass die kohärente Terminologie im ganzen Text (Projekt) verwendet wird. Solche Termbanken, die mit einem CAT-Tool integriert sind, sind selbstverständlich nicht die einzigen, die zur Verfügung gestellt werden. Es gibt auch viele andere, alleinstehende Termbanken, die unabhängig von einer Übersetzungssoftware arbeiten. Abgesehen davon, wie eine Termbank funktioniert, ohne oder mit einem CAT-Tool, soll sie entsprechend gepflegt werden. Diesbezüglich ist zu betonen, dass das Erstellen einer Termbank verlangt, nicht nur neue Termini zu sammeln und mit entsprechenden Äquivalenten in der Zielsprache zu versehen, sondern es geht auch z. B. um das Hinzufügen der hilfreichen Metadaten oder um die fortlaufende Aktualisierung einer Termbank und darum, wie diese Aktualisierung und die nachfolgende Verwaltung vorgenommen wird. Alle diese Faktoren haben einen wesentlichen Einfluss darauf, ob eine Termbank als qualitativ hochwertig bezeichnet werden darf.

TERMINOLOGIEDATENBANKEN UND IHRE QUALITÄT

Wie bereits erwähnt, bedeutet die effiziente Terminologieverwaltung vor allem Zeitersparnis durch Optimierung von Arbeitsprozessen, Beseitigung von Unklarheiten und eventuellen Missverständnissen, einheitliche Unternehmenskommunikation, Kosteneinsparungen und letztlich Kundenzufriedenheit. Der Bedarf an einer effizienten und zugleich effektiven Terminologieverwaltung ist vor allem auf die immer größeren terminologischen Ressourcen zurückzuführen. Die terminologischen Verwaltungsmaßnahmen haben zum Ziel, die ganze terminologische Infrastruktur in Ordnung zu bringen, so dass sie nicht durcheinander gerät und so die terminologische Arbeit tatsächlich erleichtern kann. Einer der wichtigsten Bestandteile der terminologischen Infrastruktur sind die erwähnten Terminologiedatenbanken (aus dem Englischen als TDBs – *Terminology Databases* bezeichnet). Die Terminologieverwaltung endet jedoch nicht mit dem Aufbau einer Terminologiedatenbank. Zur laufenden Aktualisierung und Verwaltung einer Datenbank verwendet man immer häufiger die neueste Softwaretechnologie. Diese Software wird oft mit einem Übersetzungstool verbunden, damit Übersetzer und Projektmanager bei der Bearbeitung ihrer Texte direkt auf bestimmte Informationen zurückgreifen können.

Bevor man aber mit der Terminologieverwaltung anfängt, muss man erst die Termbanken aufbauen und entsprechend mit Inhalt (terminologischen Daten – Rekorden) füllen. Wenn diese richtig konstruiert, mit dem angemessenen Inhalt gefüllt und (was genauso wichtig ist) fortlaufend aktualisiert werden, unterstützen und beeinflussen sie weitgehend die Fachübersetzungsprozesse, spiegeln den Fortschritt in Bezug auf die Harmonisierung und Normung der Termini wider und tragen schließlich zur Qualitätssicherung der Übersetzungsprozesse bei. Daher ist es notwendig, dass die Termbanken die höchste Qualität vorweisen. Ihre Qualität wird gewissermaßen durch den übersetzten Text und seine Qualität wiedergespiegelt. Darüber hinaus tragen die richtig erstellten und gepflegten (aktualisierten) Terminologiedatenbanken auch dazu bei, die Übersetzungskosten zu senken und Übersetzungsfehler zu vermeiden, indem alle Mitglieder des Projekts zur kohärenten Terminologie Zugang haben und sie anwenden.

In diesem Beitrag wird versucht, den Ausdruck Qualität in Bezug auf Terminologiedatenbanken zu definieren. Darüber hinaus wird auch der Versuch unternommen, ihre Auswertungskriterien zu bestimmen. Um das Ziel zu erreichen wird eine vergleichende Analyse von zwei Terminologiedatenbanken durchgeführt, deren Ergebnisse auf Vor- und Nachteile der heutigen terminologischen Infrastruktur hinweisen. Die erwähnte Analyse soll aber vor allem auf potenzielle Mängel der Terminologiedatenbanken und dadurch der ganzen terminologischen Infrastruktur aufmerksam machen und schließlich soll sie auch ermöglichen, die Anforderungen an zukünftige Termbanken zu formulieren. Obwohl das Thema der Termbanken nicht neu ist, wird ihm in der Fachliteratur zu wenig Beachtung geschenkt, zumindest

in Bezug auf die praktische Seite dieser Tools und auf die Möglichkeiten ihres Einsatzes im Übersetzungsprozess (mehr dazu: Sztuk 2018).

Auf zahlreichen Internetseiten zum Thema Übersetzungssoftware wird dagegen mehrmals die Rolle der Terminologiemanagementtools (oder Terminologieverwaltungstools) und Termbanken betont. Eine Terminologiedatenbank wird als eine Datenbank definiert, die Terminologiedaten sowie zugehörige Informationen enthält. Die meisten Terminologiedatenbanken sind mehrsprachig angelegt und enthalten Terminologiedaten für verschiedene Sprachen. Terminologiedaten werden mithilfe von Termbankeinträgen geordnet: Jeder Eintrag enthält mindestens einen Terminus. Ein typischer Termbankeintrag enthält beispielsweise einen Terminus, dessen Übersetzung und bestimmte Metadaten. Termbankeinträge können klassifiziert werden und jede Kategorie von Einträgen wird als Eintragsklasse bezeichnet.

Heute werden relativ viele Terminologieverwaltungssysteme angeboten. Ein Beispiel ist das SDL MultiTerm. Das ist eine Terminologiemanagement-Software, die von SDL Trados mit dem Ziel Speichern und Verwalten mehrsprachiger Terminologie erstellt wurde. Die andere bekannte Software für die Terminologieverwaltung ist QTerm von MemoQ Ltd. QTerm ist ein professionelles Terminologieverwaltungssystem für Übersetzungsfirmen und Unternehmen zum Speichern, Verwalten und Freigeben von Terminologie. Wichtig dabei sind nicht nur zusätzliche, zugehörige Informationen über Termini, sondern auch die Mehrsprachigkeit der Software. Die anderen Produzenten der Terminologiemanagementtools definieren eine Termbank ganz einfach als Wörterverzeichnis, das mit Einträgen gefüllt wird.

Aufgrund der durchgeführten Recherche und meiner Beobachtungen im Bereich der Terminologieverwaltung möchte ich meine eigene, komplexe Definition einer Termbank vorschlagen:

Eine Terminologiedatenbank (Termbank oder komplexer auch als Terminologieverwaltungssysteme – TVS bezeichnet) ist eine spezielle Art von Datenbank, die der Verwaltung von Terminologie dient. Die Anwendungszwecke können von einfachen Wörterbüchern und Glossaren bis hin zu strukturierten Thesauri (Wortnetze, dessen Begriffe durch Relationen miteinander verbunden sind) reichen. Terminologiedatenbanken unterstützen sowohl redaktionelle Arbeiten als auch die Übersetzungsprozesse. Sie helfen einheitliche, eindeutige und geprüfte Fach- und Unternehmensterminologie zu verwenden. Eine Terminologiedatenbank sorgt für eine einheitliche Kommunikation z. B. innerhalb eines Unternehmens wie auch für eine konsistente Sprache. Terminologiedatenbanken sind häufig Bestandteil von Translation-Memory-Systemen. Dort schlagen sie dem Benutzer während der Übersetzung die zielsprachliche Benennungen der verwendeten Begriffe vor. Die Daten enthalten meist neben linguistischen Informationen (z. B. Wortart, Genus, Numerus) auch fachliche Meta-Informationen, Quelle sowie Kontextbeispiele. Wesentlich dabei ist, dass die Daten eine hohe Strukturierung aufweisen und sich automatisch verarbeiten lassen, weil der Mensch selbst nicht mehr imstande ist, solche Datenmengen unter Kontrolle zu haben. Von großer Bedeutung ist dabei auch der effektive Datenaustausch unter den Datenbanken (möglichst ohne Datenverlust). Die Qualität einer Terminologiedatenbank wird nicht nur von ihrer richtigen Erstellung, sondern auch von der regelmäßigen und sorgfältigen Aktualisierung stark beeinflusst.

Diese umfangreiche Definition enthält zugleich einige Kriterien, die bei der Qualitätsauswertung einer Termbank von Bedeutung sind. Die für diesen Beitrag vorgenommene Analyse wurde unter Berücksichtigung der erwähnten Kriterien durchgeführt, um zu prüfen, ob die für die Analyse gewählten Termbanken diesen Anforderungen entsprechen.

Zusammenfassend ist zu sagen, dass eine zuverlässige Termbank u.a. terminologische Einträge zusammen mit den relevanten Metadaten enthalten soll. Die gesammelten und klassifizierten Einträge sollen editierbar sein. Eine Termbank soll den Benutzern eine Schnellsuche ermöglichen. Am besten wäre es auch, wenn die Datenbank selbst ein Update vornehmen könnte und sich dadurch automatisch aktualisiert. Eine Terminologiedatenbank soll in Bezug auf QS auch mit einem CAT-Tool im Einklang stehen, damit die in einer Termbank enthaltenen und genehmigten Termini als Übersetzungsvorschläge erkannt und beim Übersetzen angezeigt werden. Von großem Vorteil ist zweifellos, wenn der Benutzer des Tools auch die Beziehungen zwischen den einzelnen Termini sehen kann. So kann festgestellt werden, wo sich ein bestimmter Terminus in einem Termnetz (Thesaurus) befindet und in welchem Verhältnis er zu den anderen Termini steht. Am besten wäre es auch, wenn in dem Projekt ein zusätzliches Modul vorhanden wäre, das die Korrektheit und Kohärenz der verwendeten Terminologie in der Endphase der Übersetzung prüft. Diese Merkmale einer Termbank sollen deren hohe Qualität sichern.

Die erwähnte vergleichende Analyse wurde auf der Basis der erwähnten Merkmale durchgeführt. Das Hauptziel der Analyse lag darin festzustellen, ob diese Merkmale tatsächlich die Qualität (Zuverlässigkeit) einer Termbank beeinflussen bzw. garantieren.

VERGLEICHENDE, QUALITATIVE ANALYSE – VORGEHENSWEISE UND BEWERTUNGSKRITERIEN

Für die Analyse habe ich zwei umfangreiche (mit vielen Einträgen), mehrsprachige Terminologiedatenbanken ausgewählt, die sich deutlich voneinander unterscheiden. Die erste Termbank ist IATE – *Interactive Terminology for Europe*. Diese Datenbank ist für alle Benutzer jederzeit kostenfrei online zugänglich. Sie war ursprünglich als Terminologiedatenbank für die Institutionen der Europäischen Union gedacht. Die IATE-Datenbank vereint die bisher separat gepflegten Terminologiebestände verschiedener EU-Übersetzungsdienste in einer gemeinsamen Datenbank, die zur Zeit rund 1,4 Mio. Einträge enthält. Diese Datenbank kann man eher als traditionell beschreiben. Auf der Startseite kann man die Ausgangs- und die Zielsprache einstellen und zusätzlich die Suche optional auf ein bestimmtes Fachgebiet einschränken.

Durch die kleinen Sternchen wird die Zuverlässigkeit eines Terminus bestimmt. Interessant dabei ist, dass die Autoren angeben, dass jeden Tag neue Termini

eingetragen werden und der Inhalt fortlaufend aktualisiert wird. IATE trägt wesentlich dazu bei, die Qualität der schriftlichen Kommunikation der Organe und Institutionen der EU zu sichern – so die Autoren. Bei den ausgewählten Termini ist es auch möglich den Terminus im Kontext zu lesen und zu erfahren, woher er stammt (aus welchem EU-Dokument). Zusammenfassend lässt sich in Bezug auf IATE Folgendes feststellen: Die Funktionen dieser Termbank wurden ausschließlich auf die Suche nach Äquivalenten in der Zielsprache und auf die Klassifizierung der vorhandenen Termini nach den Fachgebieten beschränkt⁵.



The screenshot shows the IATE search interface. At the top left is the IATE logo with the text 'InterActive Terminology for Europe'. A language dropdown menu is set to 'Deutsch (de)'. Below the logo are links for 'Meine Suchpräferenzen', 'Suchpräferenzen löschen', and 'Hilfe'. The main search area is divided into two sections: 'Suchkriterien' and 'Optionale Kriterien'. In 'Suchkriterien', there is a search input field, a 'Suchen' button, and a dropdown for 'Ausgangssprache' set to 'bg - Bulgarisch'. Below this is a grid of checkboxes for 'Zielsprachen' including bg, cs, da, de, el, en, es, et, fi, fr, ga, hr, hu, it, la, lt, lv, mt, nl, pl, pt, ro, sk, sl, and sv. There are also links for 'Präferenzen laden' and 'Alle Löschen'. The 'Optionale Kriterien' section includes a dropdown for 'Wählen Sie das zu Ihrer Abfrage gehörende Fachgebiet' and radio buttons for 'Art der Suche' with options 'Terminus', 'Abkürzung', and 'Alles' (selected).

Abb. 1. IATE Startseite



The screenshot shows the search results for the query 'przetwarzanie danych'. The search input field contains the query and a 'Suchen' button. Below the input field, the results are displayed as a table with columns for language, subject area, and quality indicators. The results are grouped by subject area: 'Informationsverarbeitung, Datenverarbeitung, Forschung [COM]' and 'Informatik [COM]'. Each result row includes a language code (PL or DE), the query, a quality indicator (stars and a plus sign), and a 'Vollständiger Eintrag' link.

Informationsverarbeitung, Datenverarbeitung, Forschung [COM]		Vollständiger Eintrag
PL	przetwarzanie danych	★★★★+ ⓘ
DV		★★★★ ⓘ
DE	Datenverarbeitung	★★★★+ ⓘ
	Informationsverarbeitung	★★★★+ ⓘ
	Verarbeitung von Informationen	★★★★+ ⓘ
Informatik [COM]		Vollständiger Eintrag
PL	przetwarzanie danych	★★★★+ ⓘ
	Datenverwaltung	★★★★+ ⓘ
DE	Umgehen mit Daten	★★★★+ ⓘ
	Behandeln der Daten	★★★★+ ⓘ

Abb. 2. IATE Suchergebnisse

⁵ Während der Verfassung dieses Beitrags wurde das Layout der IATE-Internetseite geändert <https://iate.europa.eu/home>. Manche Funktionen wurden ergänzt, z. B. sind die Definitionen der meisten Begriffe und Informationen darüber, woher sie stammen, jetzt vorhanden. Alle Termini wurden mit Kontextbeispielen versehen.

Die andere für meine Analyse ausgewählte Terminologiedatenbank unterscheidet sich von IATE. Sie heißt COREON und ist wegen der vielen zusätzlichen Funktionen, die das System den Benutzern in Bezug auf die Terminologie und die Terminologieverwaltung anbietet, eher als ein TVS zu bezeichnen.

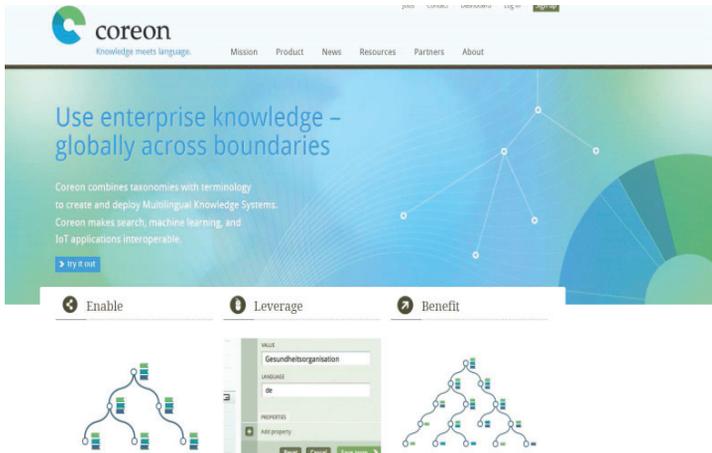


Abb. 3. COREON Startseite

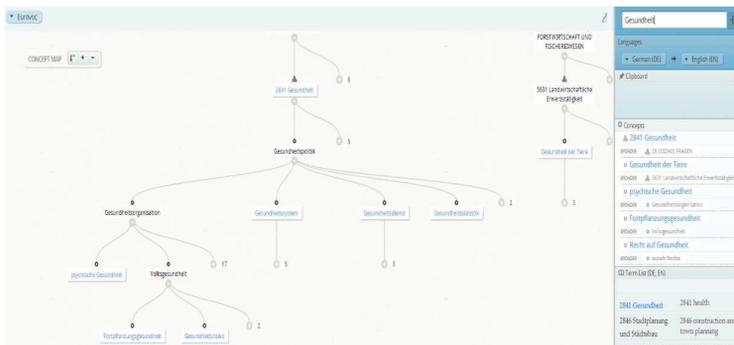


Abb. 4. COREON Suchergebnisse

Diese Bank wurde von ihren Herstellern als ein modernes *Multilingual Knowledge System* bezeichnet. Sie ist nur teilweise kostenfrei und funktioniert als eine intelligente App, die nicht nur zahlreiche Termini aus vielen Bereichen verwaltet, sondern auch die Taxonomie der Termini widerspiegelt. Dies wird durch die sog. Konzeptbäume erreicht. Die Beziehungen zwischen den einzelnen Termini werden klar dargestellt, wodurch man sofort weiß, wohin der bestimmte Terminus gehört und welcher terminologischen Gruppe er angehört. Man sieht den Oberbegriff und die Unterbegriffe. Der Beginn des Baumes ist der Oberbegriff (auch Wurzel genannt), der alle weiteren Wurzelknoten (Untertermini) enthält. Das

Merkmal aller Untertermini ist, dass sie mehr detaillierte und spezifische Angaben über das Objekt angeben als die Oberbegriffe. Diese Applikation basiert auf den neuesten Technologien aus der IT-Branche, u.a. auf der Künstlichen Intelligenz (KI)⁶. Diese Datenbank ist ebenfalls mehrsprachig, zudem kann man die Termini schnell aussuchen und sie in dem bestimmten Kontext lokalisieren.

Die der vorgenommenen Analyse zugrundeliegenden Kriterien sind:

- die Kompatibilität einer Datenbank zu anderen Datenbanken (beim Datenaustausch);
- richtige, präzise, eindeutige Äquivalente in der Zielsprache;
- die Aktualisierung der Banken (ob regelmäßig aktualisiert; mit/ ohne IT-Unterstützung);
- der Kontextbezug der angegebenen Termini, bzw. die Darstellung der Taxonomie.

Die Analyse wurde auf einer Probe von 50 Termini aus dem Bereich Datenschutz durchgeführt (sowohl Deutsch – Polnisch, als auch Polnisch – Deutsch).

ERGEBNISSE DER ANALYSE.

DEFIZITE DER TERMINOLOGISCHEN INFRASTRUKTUR

Die durchgeführte Analyse hat erwiesen, dass sich die beiden Terminologiedatenbanken tatsächlich deutlich voneinander unterscheiden. Die erste in Betracht gezogene Datenbank (IATE), die von mir als eher traditionelle Datenbank bezeichnet wurde, weil sie ohne IT-Unterstützung aktualisiert wird, hat sich im Vergleich zu der zweiten (COREON), die nach den neuesten Methoden erstellt und aktualisiert wird, als schlechter erwiesen. Unter den neuesten Methoden wird die Automatisierung und intelligente Vorgehensweise verstanden, wie etwa der Einsatz der Künstlichen Intelligenz, die heute in vielen Bereichen Triumphe feiert (mehr dazu: Sztuk 2018a, 2018b). Die Tabelle (siehe unten) veranschaulicht die Ergebnisse der Analyse:

	IATE	COREON
Kompatibilität	-	+
Richtige Äquivalente	+/-	+
Aktualisierung	-	+
Kontextbezug vs. Taxonomie	+/-	+

Abb. 5. Ergebnisse der vergleichenden Analyse

⁶ Künstliche Intelligenz (KI, auch Artifizielle Intelligenz (AI bzw. A. I.), englisch *Artificial Intelligence*, AI) ist ein Teilgebiet der Informatik, welches sich mit der Automatisierung intelligenten Verhaltens und dem Maschinellen Lernen befasst.

Wie sich herausgestellt hat, ist IATE mit den anderen Datenbanken nicht kompatibel, weil bei dem Versuch des Datenaustausches zwischen den Datenbanken ziemlich viele Daten verloren gegangen sind. Dies passiert, da die Daten in verschiedenen Formaten gespeichert wurden. Solche Datenverluste sind nicht nur unerwünscht, sondern auch kostenintensiv. COREON dagegen zeichnet sich durch einen sehr strukturierten Aufbau aus. Alle Daten sind nach dem gleichen (standardisierten), informatischen Muster eingetragen (gespeichert). Das ermöglicht den verlustfreien Datenaustausch.

In Bezug auf das zweite Kriterium, also die richtigen, präzisen und eindeutigen Äquivalente kann man bei IATE nicht eindeutig feststellen, ob diese Datenbank effektiv ist. Die Analyse hat nämlich erwiesen, dass diese Datenbank zwar richtige Äquivalente anbietet, aber nur in ca. 75% aller Anfragen. In den übrigen Fällen bietet die Datenbank entweder keine Äquivalente an, oder im Gegenteil zu viele, was im Fall der Termini unerwünscht ist, weil Termini präzise und eindeutig sein sollen. Synonymie ist bei der Terminologie auf jeden Fall unerwünscht. Bei COREON dagegen habe ich in 50 Anfragen 47 richtige, präzise Äquivalente gefunden, was rund 94% ausmacht.

Bei dem dritten Kriterium, das ich Aktualisierung genannt habe, ging es darum festzustellen, ob und wie die Datenbank aktualisiert wird. Für diesen Zweck habe ich 15 Termini gewählt, die sich auf die neuesten Methoden des Missbrauchs personenbezogener Daten beziehen. Bei IATE habe ich keine Einträge gefunden. Bei COREON habe ich 14 Äquivalente von den 15 ausgewählten Termini gefunden. Was die Art und Weise der Aktualisierung angeht, wird IATE manuell gepflegt, also von Menschen, die die Textkorpora durchsuchen und neue Termkandidaten für die Bank vorschlagen. Diese Methode ist recht Zeit- und Kostenaufwendig. Bei COREON werden Textkorpora automatisch durchsucht, um neue Termkandidaten vorzuschlagen. Der Nutzer muss dann nur verifizieren, ob die Ergebnisse zuverlässig sind und in die Datenbank einbezogen werden können.

Das letzte in Betracht gezogene Kriterium, habe ich Kontextbezug vs. Taxonomie genannt. Bei IATE gibt es bei allen Termini Beispielsätze (Kontext), aber man weiß nicht, wo sich der bestimmte Terminus in dem terminologischen Netz (Thesaurus) befindet. Dadurch sind dem Nutzer auch die Beziehungen unter den Termini nicht bekannt. Bei COREON sieht es anders aus: Termini sind in Form von Konzeptkarten dargestellt und die Beispielsätze sind ebenfalls immer vorhanden. So sieht man sofort, dass z. B. Tiger der Katzenfamilie angehört, wobei der Tiger, als ein Unterterminus zu der Katze, mit mehr spezifischen Merkmalen versehen wurde.

QUALITÄT IN BEZUG AUF TERMINOLOGIEDATENBANKEN. TERMBANKEN DER ZUKUNFT – ANFORDERUNGEN

Die theoretischen Erörterungen und die durchgeführte Analyse haben ermöglicht, mit einer ganz anderen Perspektive auf die Terminologiedatenbanken und deren Qualität zu schauen und bestimmte, wesentliche Schlussfolgerungen bezüglich ihrer Qualität zu ziehen, die zugleich als Anforderungen an zuverlässige Termbanken der Zukunft interpretiert werden können. Die richtige Erstellung einer Terminologiedatenbank ist genauso wichtig wie ihre regelmäßige Aktualisierung, die am besten mithilfe der neuesten, automatisierten IT-Methoden erfolgt. Man kann davon ausgehen, dass die IATE Datenbank, die nicht automatisch aktualisiert wird, deswegen höchstwahrscheinlich nicht so zuverlässig ist wie z. B. die COREON Datenbank, die teilweise automatisch aktualisiert wird, was einen deutlichen Beitrag zu ihrer besseren Qualität leistet.

Terminologiedatenbanken, die nicht mit den anderen Datenbanken kompatibel sind, haben meines Erachtens keine Zukunft. Heutige Wirtschaftssubjekte wollen immer häufiger ihre Daten ‚reibungslos‘ austauschen. Wenn eine Datenbank nicht zu den anderen passt (wegen unterschiedlicher Datenspeicherformate), kommt es zum kostspieligen Datenverlust oder zur Datenverformung (auch Datenverfälschung genannt). Es stellt sich also die Frage, wie Terminologiedatenbanken erstellt werden sollen, damit sie miteinander kompatibel sind. Hier kommt die ISO Norm TBX (*TermBase eXchange*) 30042 zur Hilfe. Diese Norm gibt Richtlinien an, wie Terminologiedatenbanken erstellt werden, wie terminologische Einträge aussehen und wie sie gespeichert werden sollen, damit die terminologischen Daten problemlos ausgetauscht werden.

Konzeptkarten oder Konzeptbäume haben sich als sehr nützlich herausgestellt. Sie stellen deutlich dar, wo ein Terminus hingehört, in welchen Beziehungen er zu den anderen Termini von der jeweiligen Familie steht. Das kann äußerst nützlich sein, z. B. für einen Fachübersetzer. Es ist an dieser Stelle hinzuzufügen, dass diese Konzeptbäume bei COREON auch (teilweise) automatisch geschaffen werden. Die Nutzer überwachen diese Abläufe lediglich und korrigieren sie, falls nötig.

Dabei lassen sich einige allgemeine Tendenzen beobachten, u.a., dass die Automatisierung der Erstellungs- und Aktualisierungsprozesse von Terminologiedatenbanken sehr stark ihre Qualität beeinflusst. Menschen sind nicht imstande große Datenmengen unter Kontrolle zu haben, deswegen sind Maschinen, also Computer unentbehrlich, weil sie jede Menge Daten verarbeiten können. Die menschliche Rolle wird dabei sehr eingeschränkt, nämlich nur zur Überwachung, ob alles richtig verläuft.

Aufgrund meiner Untersuchungen lässt sich die folgende Definition von Qualität in Bezug auf Terminologiedatenbanken formulieren: Eine zuverlässige Termbank soll die Summe bestimmter Anforderungen, sowohl in Bezug auf die Erstellung als auch auf die Aktualisierung, erfüllen. Diese Anforderungen

entsprechen zugleich gewissermaßen den von mir zur Analyse ausgewählten Kriterien der Qualitätsauswertung. Diesbezüglich lässt sich schlussfolgern, dass die Qualität einer Terminologiedatenbank umso besser ist, je mehr Auswertungskriterien erfüllt werden. Zusammenfassend kann man sagen, dass eine Terminologiedatenbank von guter Qualität ist, wenn sie: (a) mit anderen Terminologiedatenbanken kompatibel ist; (b) richtige, präzise und eindeutige Äquivalente anbietet; (c) regelmäßig aktualisiert wird (bei mehreren terminologischen Einträgen am besten, wenn der Aktualisierungsprozess mindestens teilweise automatisch verläuft); (d) die Kontextbezüge und Taxonomie der Termini angibt, so dass die Beziehungen zwischen den einzelnen Einträgen transparent präsentiert werden.

BIBLIOGRAPHIE

- GAJDA, S. (1990): *Wprowadzenie do teorii terminu*, Opole.
- GRUCZA, F. (red.) (1991): *Teoretyczne podstawy terminologii*, Wrocław.
- GRUCZA, F. (2002): „Języki specjalistyczne – indykatory i/ lub determinanty rozwoju cywilizacyjnego”, in: J. LUKSZYN (ed.): *Języki specjalistyczne. Problemy technolingwistyki*, Warszawa.
- GRUCZA, F. (2008): „O językach dotyczących europejskiej integracji, Unii Europejskiej i potrzebie ukonstytuowania ogólnej lingwistyki języków specjalistycznych”, in: *Podstawy technolingwistyki I*, Warszawa, s. 27–74.
- GRUCZA, S. (2002): „Badania nad językami specjalistycznymi w Niemczech”, in: *Języki specjalistyczne. Problemy technolingwistyki*, Warszawa, s. 81–100.
- GRUCZA, S. (2004): *Od lingwistyki tekstu do lingwistyki tekstu specjalistycznego*, Warszawa.
- GRUCZA, S. (2008): *Lingwistyka języków specjalistycznych*, Warszawa.
- GRUCZA, S. (2010): „Język a poznanie – kilka uwag na temat rozważania ich wzajemnych relacji”, in: S. GRUCZA/ A. MARCHWIŃSKI/ M. PŁUZYCZKA (ed.): *Translatoryka – koncepcje, modele, analizy*, Warszawa, s. 54–68.
- JURKOWSKI, M. (1991): „Metajęzyk terminologii”, in: F. GRUCZA (ed.): *Teoretyczne podstawy terminologii*, Warszawa, s. 45–60.
- KOCKAERT, H.J./ STEURS, F. (2015): *Handbook of Terminology*, Vol. 1, Amsterdam.
- LUKSZYN, J. (1991): „Lingwistyczne problemy badań terminologicznych”, in: *Teoretyczne podstawy terminologii*, Warszawa–Wrocław–Kraków, s. 79–96.
- LUKSZYN, J. (1998): *Tezaurus terminologii translatorycznej*, Warszawa.
- LUKSZYN, J./ ZMARZER, W. (2001): *Teoretyczne podstawy terminologii*, Warszawa.
- NOWICKI, W. (1996): *Podstawy terminologii*, Wrocław.
- POPIOŁEK, M. (2015): “Terminology Management within a Translation Quality Assurance Process”, in: KOCKAERT, H.J./ STEURS, F.: *Handbook of Terminology*, Vol. 1, Amsterdam.
- SZTUK, A. (2017): *Normalizacja translatoryczna*, Warszawa.
- SZTUK, A. (2018a): „Sztuczna inteligencja (SI) w badaniach lingwistycznych”, *Applied Linguistics Papers*, 25/4, Warszawa, s. 159–174.
- SZTUK, A. (2018b): „Nowoczesne metody aktualizacji baz danych terminologicznych. Statystyka za cenę prywatności?”, in: *Komunikacja Specjalistyczna*, 14/2018, Warszawa, s. 127–138.
- WRIGHT, S.E./ BUDIN, G. (ed.) (1997): *Handbook of Terminology Management*, Vol. 1: Basic Aspects of Terminology Management, Amsterdam/Philadelphia.
- WRIGHT, S.E./ BUDIN, G. (ed.) (2001): *Handbook of Terminology Management*, Vol. 2: Application-Oriented Terminology Management, Amsterdam–Philadelphia.

WÜSTER, E. (1979): *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*, Band 8, Teil 1, Bonn.

NORMEN

ISO 704 *Terminology Work – Principles and Methods*.

ISO 12616 *Translation-oriented Terminography*.

ISO 12620 *Terminology and other language and content resources – specification of data categories and management of a data category registry for language resources*.

ISO 30042 *Systems to manage terminology, knowledge and content – TermBase eXchange (TBX)*.

ONLINEQUELLEN

<https://de.wikipedia.org/wiki/Qualit%C3%A4tssicherung> (Zugriff: 23.08.2019).

<https://iate.europa.eu/home> (Zugriff: 15.08.2019).

<https://www.coreon.com> (Zugriff: 13.08.2019).