Deutscher Terminologie-Tag e.V.

21. Jahrgang | www.dttev.org

edition

Fachzeitschrift für Terminologie

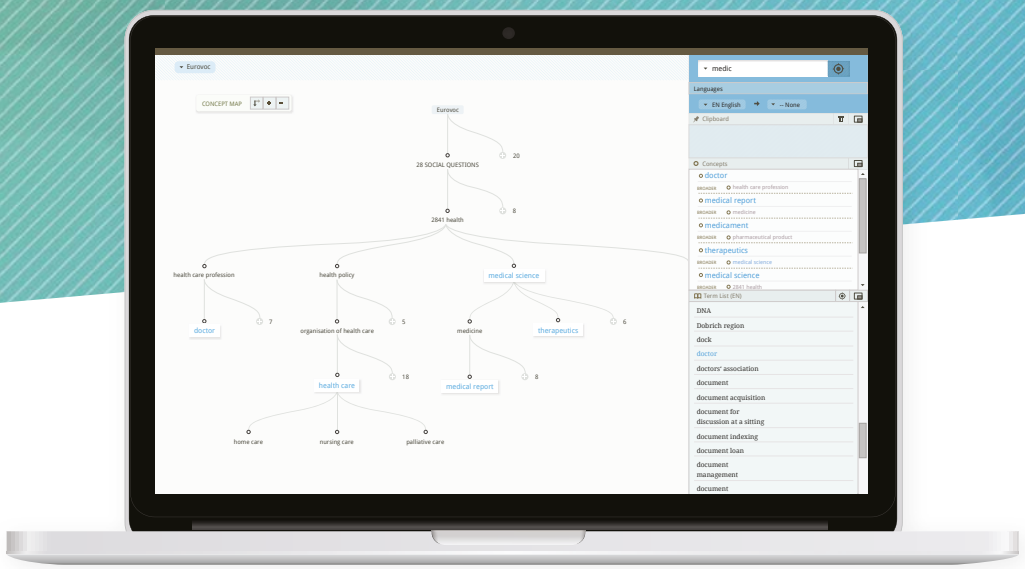
2|25

Terminologie und KI

eine smarte Verbindung

Your Multilingual Knowledge System

Extreme Visual Terminology



360-degree control through systematic approach: **multilingual Concept Maps**

Proven enterprise deployments, scalable with Single sign-on: appealing, **plug-in free browser solution**

Enable NMT workflows, tune enterprise search, annotate and classify texts: **versatile integrations**

coreon

Knowledge meets language.

coreon.com

„Terminologie und KI – eine smarte Verbindung“

Was vor wenigen Jahren noch visionär klang, gehört heute längst zum Arbeitsalltag. Gleichzeitig stehen wir mitten in einer spannenden Phase – zwischen technologischer Begeisterung und methodischer Sorgfalt, zwischen Automatisierung und fachlicher Verantwortung. Genau in diesem Spannungsfeld bewegt sich die vorliegende Ausgabe.

Mo Han und Marcus Müller widmen sich terminologischen Innovationen in den Internationalen Beziehungen und zeigen, wie große Sprachmodelle zur Disambiguierung von Termini im politikwissenschaftlichen Forschungsbereich der Internationalen Beziehungen beitragen können. Sinan Sen führt uns in die Welt der semantischen Suche und erklärbaren KI und skizziert neue Wege in der Literaturrecherche. Vanessa Šorak beleuchtet die Terminologieerzwingung mit generativen LLMs – hier geht es um Konsistenz in Zeiten von Textproduktion auf Knopfdruck. Maurice Mayer-Dewor und Giovanni Rovere stellen schließlich die Frage: Brauchen wir professionelle Übersetzungswörterbücher eigentlich noch? Sie zeigen, warum Valenz- und Fachwörterbücher mehr leisten als reine Wortlisten zur Verfügung zu stellen.

Darüber hinaus stellt Petra Drewer die in Arbeit befindliche DIN 19460 vor, Tom Winter berichtet von der NFDI4ING-Konferenz 2025 und Angelika Ottmann hat das Konnektorenwörterbuch des IDS ausprobiert.

Die Beiträge zeigen einmal mehr, dass Terminologie im KI-Zeitalter eine Schlüsselposition einnimmt.

Wir wünschen Ihnen eine anregende Lektüre!

Herzlichst
Ihre Redaktion

geschrieben unter Mitwirkung von ChatGPT

Dr. Nicole Keller
Redaktionsleitung
redaktion@dttev.org

Angelika Ottmann
Redaktionsleitung
redaktion@dttev.org

Inhalt

21. Jahrgang.
edition erscheint zweimal im Jahr (Juni/Dezember).
ISSN 1862-023X.
Herausgeber: Deutscher Terminologie-Tag e.V. (DTT) – www.dttev.org
Redaktionsleitung: Dr. Nicole Keller und Angelika Ottmann – redaktion@dttev.org
Titelseite, DTT-Anzeigen, Layout und Satz: Tamara Arndt – layout@dttev.org
Anzeigen: Olga Buchstaller-Vodopiyanova – anzeigen@dttev.org
Lektorat: Maurice Mayer-Dewor
Mediadaten und vollständiges Impressum: www.dttev.org/edition/impressum

Terminologische Innovationen in Internationalen Beziehungen

Mo Han und Marcus Müller

The DFG funded project "Terminological Innovations in International Relations (IR)" aims at a systematic investigation of term usages across different domains of German discourse about IR from 1976 to 2000. Due to the ambiguous characteristics of IR terminology, LLM-based approach has been applied to identify their contextual sense and shown a promising outcome.

Keywords: terminology, corpus linguistics, international relations, word sense disambiguation, LLM

Obwohl unter Terminologie häufig eine Menge an Benennungen verstanden wird, die sich auf eindeutige Weise auf einen Begriff beziehen, zeigt sich in der Praxis, dass Termini je nach Kontext durchaus mehrdeutig sein können. Dabei gilt die erst auf den zweiten Blick intuitive Regel, dass die Mehrdeutigkeit von Termini mit dem Abstraktionsgrad ihres Anwendungsfeldes steigt: Während die Bezeichnung von Nägeln im Baumarkt in der Regel dem Ideal der Fachterminologie entspricht, sind Termini in der Wissenschaft graduell vage und oft auch polysem. Das gilt durchaus auch für vermeintlich „harte" Wissenschaften wie die Biologie [23], umso mehr aber für die Geistes- und Sozialwissenschaften. Bei der Identifizierung und Extraktion von Termini stellen sich hier besondere Herausforderungen, deren Überwindung aber auch Erkenntnis für die Terminologieextraktion in technischen Praxisfeldern verspricht, frei nach Frank Sinatras Motto: If I can make it there, I'll make it anywhere.

Unser Beitrag widmet sich daher der Disambiguierung von Termini im politikwissenschaftlichen Forschungsbereich der Internationalen Beziehungen (IB), in dem abstrakte Theoriebildung und praktische Politik gleichermaßen verhandelt werden, was die Polysemie zentraler Termini unterstützt. Nach einer Darstellung des übergeordneten Forschungsprojekts wird im weiteren Verlauf eine Pilotstudie vorgestellt.

Das TIIB-Projekt

Das Reden und Schreiben über Internationale Beziehungen ist geprägt von abstrakten, oft metaphorischen Termini wie *internationales System*, *Gleichgewicht der Mächte* oder *internationale Anarchie*. Viele dieser Termini wurden in einem fachsprachlichen Diskurs entwickelt oder dort mit einer neuen, vom alltäglichen Sprachgebrauch abweichenden Bedeutung versehen. Da die bisherigen Forschungen zu Begriffs- und Diskursgeschichte sich meist auf die Analyse und Interpretation einiger Schlüsselwerke [24] stützen, die den weiteren IB-Diskurs abbildet, hat sich das hier vorgestellte Projekt „Terminologische Innovationen in Internationalen Beziehungen" (abgekürzt: TIIB) vorgenommen, auf einer breiten Datenbasis den terminologischen Wandel des akademischen IB-Diskurses nachzuzeichnen, und zwar auch in seinen möglichen Auswirkungen auf die politische Praxis[1].

Vor diesem Hintergrund geht es uns um zwei Forschungsfragen:

1. Auf welchen Wegen und in welchem Ausmaß gehen terminologische Innovationen[2] vom wissenschaftlichen Fachdiskurs in die politische Alltagssprache ein und vice versa?
2. Welche möglichen Auswirkungen haben die Begriffsübernahmen auf die politische Praxis und welche Schule der IB ist dabei mit ihrer spezifischen Terminologie besonders erfolgreich?

[1] Wenn politikwissenschaftliche Termini in die politische Praxis übernommen werden, können sie konkreten Einfluss auf die politische Realität nehmen, indem Sprache auch Ideen, Normen sowie Werte vermitteln kann und eine konstitutive Rolle für den Gegenstand der Internationalen Beziehungen spielt (vgl. den sozialkonstruktivistischen Ansatz in den Internationalen Beziehungen, grundlegend Onuf 1989, Wendt 1999 [17,25]).

[2] Unter Innovation wird dabei im Sinne der soziopragmatischen Sprachwandelforschung eine lexikalische Neuerung verstanden, die als Reaktion auf eine durch lebensweltliche, ideengeschichtliche oder sprachsystematische Entwicklungen entstandene Bezeichnungslücke in der Sprache entsteht.

Das TIIB-Korpus

Methodisch bedingt das Vorhaben den Einsatz korpuslinguistischer Mittel. Dafür haben wir ein dreiteiliges Korpus[3] angelegt, das aus folgenden Subkorpora besteht: (1) Akademischer Diskurs: Fachzeitschriften zu IB[4]; (2) Politikberatung: Berichte von Think-Tanks und der Beratungsstelle des Bundestags; (3) Politische Praxis: Plenarprotokolle, parlamentarische Antworten sowie Ministerreden.

Bisher liegen die Subkorpora „Akademischer Diskurs" (5.425 Texte, 40.768.562 Tokens; 2.294 Texte, 13.236.245 Tokens) und „Politische Praxis" (21.941 Texte, 12.676.872 Tokens) vor.

Terminologie im Fach Internationale Beziehungen

Komplementär haben wir eine online-Befragung von ca. 30 deutschsprachigen IB-Expert*innen mit der Bitte um Nennung relevanter terminologischer Innovationen in der Disziplin durchgeführt, die uns als zusätzliche Suchausdrücke dienen. In der Expertenliste stehen mehr als 200 Kandidatentermini, von denen die meisten mehrdeutig sind.

Das mag nicht überraschend sein, denn Terminologie ist in den meisten Sozialwissenschaften semantisch komplex, da Konzepte häufig Phänomenkomplexen entsprechen und nicht Einzelerscheinungen [14]. Verschiedene Denkschulen und theoretische Ansätze prägen ihre eigene Terminologie, die häufig normative oder ideologische Konnotationen trägt, obwohl dieselben empirischen Probleme und Phänomene diskutiert werden [13,21]. Termini sind daher oft nicht explizit definiert und standardisiert [11]. Wie Schnedermann [20] betont, sind Gebrauch und Verständnis der Terminologie kontextabhängig, weshalb bei der Analyse terminologischer Praktiken die spezifischen situativen Nuancen berücksichtigt werden müssen.

Ein illustratives Beispiel stellt der Terminus *Regime* dar. Dieses aus dem Französischen entlehnte deutsche Wort geht auf das lateinische „regimen" in der Bedeutung „Kontrolle" und „Leitung" zurück. In der Alltagssprache ist der Ausdruck negativ konnotiert, er wird zur Bezeichnung demokratisch zweifelhafter, nicht legitimierter oder totalitärer Regierungen verwendet [19]. Im Kontext der Internationalen Beziehungen hingegen wird *Regime* durch eine explizite Definition kontrolliert. Als Standardreferenz dient die von Krasner [8] formulierte Begriffsbestimmung: Ein Regime ist demnach ein Komplex aus "implicit or explicit principles, norms, rules and decision-making procedures around which actors' expectations converge in a given area of international relations". Solche Regime existieren in unterschiedlichen Politikfeldern: Sie regulieren den Welthandel (GATT), sichern die Einhaltung des Atomwaffensperrvertrags (NPT), steuern die konventionelle Abrüstung in Europa (CFE-Vertrag) und schützen die arktischen Eisbären [18]. Diese beiden Bedeutungen von *Regime* – die alltagssprachliche und die fachspezifische der IB – koexistieren in der akademischen Literatur und werden nicht selten sogar innerhalb desselben Textes verwendet.

Dieser mehrdeutige Charakter der Terminologie in den IB stellt eine zentrale Herausforderung dar. Deshalb bildet die Identifikation der jeweiligen Gebrauchsbedeutung einen wesentlichen Bestandteil der vorgesehenen semantischen Annotation.

Experimente zur Identifikation der IB-relevanten Bedeutungen

Relevante Studien

Angesichts dieser methodischen Herausforderung benötigen wir eine replizierbare Methode zur Disambiguierung von Fachtermini. In einer Pilotstudie zu *Regime* wurde eine Liste möglicher Attribute, die verschiedene Bedeutungen beschreiben, ausgearbeitet [21]. Diese Herangehensweise gewährleistet zwar eine präzise Klassifikation, erweist sich jedoch als zeitaufwändig und personalintensiv bei der Übertragung auf weitere Termini.

Es stellt sich uns also eine Aufgabe im Bereich Word Sense Disambiguation (WSD). Das ist ein Prozess im Methodenrahmen des Natural Language Processing (NLP), bei dem es darum geht, einem Textwort (Token) aus einem Set vordefinierter Bedeutungen die passende entsprechend dem Kontext zuzuordnen [1,3]. In den letzten Jahren hat WSD durch wissensbasierte, überwachte und semi-überwachte Ansätze sehr gute Ergebnisse erzielt[5]. In jüngeren Studien verlagerte sich der Fokus auf generative Large Language Models (LLMs), also KI-Systeme wie GPT, die auf der Grundlage großer Textkorpora trainiert werden. Diese Modelle sind sehr gut darin, komplexe sprachliche Muster zu erfassen und kohärente Texte zu generieren. So wurden in verschiedenen NLP-Aufgaben bedeutende Erfolge erzielt [9]. LLMs haben auch großes Potenzial für WSD, obwohl einige Studien darauf hindeuten, dass dieser Ansatz ohne Finetuning die aktuellen State-of-the-Art Methoden noch nicht übertreffen kann [1,7,22,26]. Unter Finetuning versteht man das Weitertrainieren eines Sprachmodells auf einem aufgabenrelevanten Datensatz.

[3] Mehr zu den Korpusquellen sowie zum Preprocessing siehe https://www.discourselab.de/moodle/mod/resource/view.php?id=389 (Zugriff: 18.07.2025), Müller et al. (2020) [12].

[4] Dieses Teilkorpus gliedert sich intern noch einmal in internationale, englischsprachige Publikationen und deutschsprachige wissenschaftliche Publikationen zu IB, um Vergleiche zu ermöglichen.

[5] Ein Überblick siehe Loureiro et al. (2021) [10]; Bevilacqua et al. (2021) [3].

In unserem Experiment haben wir die Disambiguierung zu einer binären Klassifikationsaufgabe vereinfacht, bei der zu entscheiden ist, ob ein gegebener Ausdruck in einem Satz als IB-Term betrachtet werden kann.

Zur Illustration geben wir das Beispiel einer binären Klassifikation von *Regime*:

- IB-Term (1): Bei diesen beiden Konflikten ist durch die Etablierung internationaler Regime eine neue Ära des Konfliktaustrages eingeläutet worden.
  (Text-ID: AD_ZIB_1997_12_00_00001)
- Kein IB-Term (0): Nach dem Ende des Regimes von Siad Barre im Jahr 1991 kam es in Somalia zu chaotischen Zuständen:
  (Text-ID: AD_FW_1996_00_00_00010)

Unsere Hypothese ist, dass wir für diese binäre Klassifikationsaufgabe mit LLMs sehr gute Ergebnisse bekommen. Das möchten wir prüfen. LLMs stellen im Vergleich zu neuronalen Netzen der Vorgängergeneration insgesamt geringere Anforderungen an Trainingsdaten und Finetuning und lassen sich durch natürliche Sprache programmieren.

Für das Klassifikationsexperiment wurden sechs Termini ausgewählt, die sowohl in der Expertenliste als auch im Korpus prominent auftraten. Neben *Regime* (s.o.) wurden *Entspannung, Kooperation, Integration, Intervention* und *Norm* ausgewählt.

Dataset und Preprocessing

Die ausgewählten Termini weisen Mehrdeutigkeiten auf verschiedenen Ebenen auf: Sie können (1) unterschiedliche Bedeutungen tragen oder (2) dieselbe Bedeutung auf verschiedenen Bezugsebenen anwenden. Fall 2 wurde jeweils durch Asterisk (*) markiert. Im Folgenden wird ein Überblick über die relevanten Bedeutungen der Termini gegeben:

Entspannung bezeichnet in den IB „politische Maßnahmen, die darauf angelegt sind, Krisensituationen zu mildern und mögliche Konflikte zu vermeiden“ [19], wobei die Handlungssubjekte im IB-Kontext Staaten sind. Entspannungspolitik bezieht sich speziell auf die politischen Maßnahmen zur Regulierung des Ost-West-Konfliktes und diese Zeitperiode.

Unter ***Kooperation**** wird auf internationaler Ebene die Überwindung von Nicht-Zusammenarbeit, von Nebeneinanderarbeit isoliert handelnder Einheiten verstanden [6].

Der Terminus ***Integration**** wird in den IB verwendet für „den Prozess, einzelne Staaten in größeren Einheiten zusammenzufassen und einen neuen, übergeordneten Akteur zu bilden“ [2]. *Integration* wird in der vorliegenden Studie sowohl berücksichtigt, wenn der Terminus sich auf den Endzustand des Prozesses bezieht, als auch dann, wenn der Weg dorthin thematisiert wird.

Intervention* bezieht sich im IB-Bereich auf das (oft bewaffnete) Eingreifen eines oder mehrerer Staaten bzw. einer oder mehrerer Internationaler Organisationen in die inneren Angelegenheiten anderer Staaten [16].

Unter einer internationalen ***Norm**** werden gemeinsam geteilte Erwartungen über angemessenes Verhalten von Regierungen und bestimmten nicht-staatlichen Akteuren auf internationaler Ebene verstanden [15]. *Normen* sind demnach unverbindliche Rahmenwerke, wie z. B. freiwillige Verhaltenskodizes oder Konventionen, und können den Rahmen für formellere, verbindliche Vereinbarungen bilden.

Aus dem TIIB-Korpus wurden 195 Sätze pro Terminus zufällig als Testdatensatz extrahiert, wobei Sätze aus denselben Texten möglichst vermieden wurden. Für jeden Terminus wurden zusätzlich 5 weitere Sätze aus dem Korpus und 5 Sätze aus dem Kernkorpus des Digitalen Wörterbuchs der Deutschen Sprache (DWDS) manuell ausgewählt, um die Aufgabenbeispiele zu erstellen. Die Aufgabenbeispiele enthalten dabei die gleiche Anzahl von Sätzen mit IB-spezifischen und nicht IB-spezifischen Wortbedeutungen. Das relevante Wort im Satz wurde mittels doppeltem Asterisk hervorgehoben. Um zusätzliche Kontext-Informationen zu geben, haben wir jeweils 20 Tokens vor und nach dem Satz sowie den Titel des jeweiligen Textes beigefügt. Alle Sätze wurden von zwei Annotierenden klassifiziert, problematische Fälle wurden mit einem dritten IB-Experten diskutiert.

Gestaltung von Prompts und Modellselektion

Um zu untersuchen, inwiefern Definitionen und Aufgabenbeschreibungen die Leistung eines Modells beeinflussen, haben wir für jeden Terminus drei Prompts[6] im Rahmen des RACE Prompting Frameworks[7] entwickelt und in englischer Sprache formuliert[8].

Im **Baseline Prompt** wurde dem Modell als IB-Experte die Aufgabe gegeben, die Bedeutung des Wortes im Satz als IB-relevant oder nicht IB-relevant einzuordnen (Your task is to determine whether the target word “Regime” in the given sentence is IR-related (1) or not (0)). Auf dem Baseline Prompt bauen der Sense Definition Prompt und der Stepwise Frame Prompt auf. Wie der Name schon andeutet, enthält der **Sense Definition Prompt** Bedeutungsbeschreibungen, etwa nach dem Muster: „Regime“ wird als IB-bezogen (1) betrachtet, wenn es sich auf … bezieht, und als nicht IB-bezogen (0), wenn es sich auf … bezieht.

[6] Aus Platzgründen werden die Prompts nicht detailliert und vollständig aufgelistet. Die gesamten Prompts siehe Github Repository „TIIB-Term-Sense-Identification“ (URL: https://github.com/LucienBlue/TIIB-Term-Sense-Identification).

[7] Das RACE Prompting Framework fokussiert auf *Role*, *Action*, *Context* und *Execution*.

[8] Ein Vortest mit begrenzter Anzahl von Sätzen hat gezeigt, dass es keinen Unterschied macht, in welcher Sprache der Prompt formuliert wird.

	Baseline					**Sense Definition**					**Stepwise Frame**				
	gpt-4o-mini	gpt-4.1	gpt-o4-mini	ds-v3	ds-r1	gpt-4o-mini	gpt-4.1	gpt-o4-mini	ds-v3	ds-r1	gpt-4o-mini	gpt-4.1	gpt-o4-mini	ds-v3	ds-r1
Entspannung	97.3%	97.5%	98.4%	97.8%	97.3%	97.5%	**98.6%**	98.1%	98.4%	98.1%	93.1%	95.7%	86.6%	97.5%	95.7%
Kooperation	93.8%	94.6%	96.7%	93.5%	96.4%	94.8%	96.7%	94.0%	95.3%	96.0%	94.2%	**97.0%**	94.6%	96.4%	95.3%
Integration	88.7%	95.5%	97.8%	94.2%	**97.9%**	86.7%	87.6%	80.5%	95.5%	86.0%	84.6%	97.1%	91.8%	96.7%	90.1%
Intervention	95.9%	93.9%	97.1%	**98.1%**	95.3%	95.7%	97.0%	97.4%	97.0%	94.6%	84.2%	97.4%	91.9%	96.4%	94.2%
Norm	84.0%	86.3%	90.9%	85.8%	91.4%	88.9%	91.8%	**92.7%**	88.9%	91.2%	88.6%	88.2%	85.0%	89.7%	83.0%
Regime	88.2%	83.7%	71.8%	62.6%	65.5%	95.1%	97.4%	96.6%	97.5%	94.9%	96.6%	**99.2%**	95.7%	94.8%	98.3%

Tab. 1: F1-Werte für alle Modelle und Prompt-Varianten

	Entspannung			**Kooperation**			**Integration**		
	Baseline	Sense Definition	Stepwise Frame	Baseline	Sense Definition	Stepwise Frame	Baseline	Sense Definition	Stepwise Frame
Precision	95.4%	96.6%	97.3%	91.7%	95.6%	96.3%	90.9%	97.6%	96.4%
Recall	100.0%	99.7%	90.7%	98.7%	95.3%	95.0%	99.4%	79.4%	88.3%
F1-Wert	97.7%	**98.1%**	93.7%	95.0%	95.4%	**95.5%**	**94.8%**	87.2%	92.0%
	Intervention			**Norm**			**Regime**		
	Baseline	Sense Definition	Stepwise Frame	Baseline	Sense Definition	Stepwise Frame	Baseline	Sense Definition	Stepwise Frame
Precision	93.2%	98.5%	96.6%	81.1%	88.4%	87.0%	61.4%	97.7%	99.3%
Recall	99.1%	94.2%	89.7%	95.9%	93.5%	87.9%	96.3%	95.0%	94.7%
F1-Wert	96.0%	**96.3%**	92.8%	87.7%	**90.7%**	86.9%	74.4%	96.3%	**96.9%**

Tab. 2: Durchschnittliche Leistung aller Termini und Prompts

	gpt-4o-mini			**gpt-4.1**			**gpt-o4-mini**			**ds-v3**			**ds-r1**		
	Baseline	Sense Definition	Stepwise Frame	Baseline	Sense Definition	Stepwise Frame	Baseline	Sense Definition	Stepwise Frame	Baseline	Sense Definition	Stepwise Frame	Baseline	Sense Definition	Stepwise Frame
Precision	84.8%	91.9%	91.6%	86.3%	96.6%	95.0%	88.6%	**98.2%**	97.4%	82.2%	94.9%	95.4%	86.0%	97.1%	97.9%
Recall	99.3%	94.7%	89.5%	**98.6%**	93.6%	96.8%	97.0%	89.4%	85.3%	98.5%	96.1%	95.2%	97.7%	90.5%	88.4%
F1-Wert	91.3%	**93.1%**	90.2%	91.9%	94.9%	**95.8%**	92.1%	**93.2%**	90.9%	88.7%	**95.4%**	95.2%	90.6%	**93.5%**	92.8%

Tab. 3: Durchschnittliche Leistung aller Modelle

Der **Stepwise Frame Prompt** basiert auf der Frame-Semantik. Das ist ein Paradigma, das die Bedeutung von Wörtern über ihren Kontext bestimmt. Die Frame-Semantik geht davon aus, dass sich Sprachgebrauch in realen Situationen vollzieht und es damit eine komplexe Beziehung zwischen Sprachbedeutung und Situationswahrnehmung gibt [5]. Frames konstituieren sich als geordnete Strukturen aus Begriffs- und Wissenselementen. Strukturell betrachtet weisen Frames einen zentralen Frame-Kern auf, der als thematisches Fundament des jeweiligen Frames fungiert. Dieser Kern wird von einer systematischen Konstellation von Frame-konstituierenden Elementen bestimmt. Diese Frame-Elemente erfüllen die Funktion von Anschlussstellen, sogenannten *slots*, denen im Sprachgebrauch konkrete Konzepte als *fillers* zugeordnet werden [4]. Die zugrundeliegende Idee dafür besagt, dass distinkte Systembedeutungen jeweils spezifischen Frames entsprechen, während sich die referentielle Bedeutung eines Ausdrucks durch die kontextuelle Aktualisierung des Frames manifestiert. Daher wird der Prompt in zwei Schritte unterteilt: Der erste Schritt besteht in der Identifikation der im Prompt definierten Frames[9] und der zweite in der Identifikation von *fillers*. Drei Modelle aus der GPT-Familie von Open AI wurden über die OpenAI-API getestet: gpt-4.1-2025-04-14 (im Folgenden: gpt-4.1), gpt-4o-mini-2024-07-18 (im Folgenden: gpt-4o-mini) und o4-mini-2025-04-16 (im Folgenden: gpt-o4-mini)[10]. Zusätzlich wurden zwei Modelle von DeepSeek über DeepSeeks API getestet: deepseek-chat (deepseek-v3-0325, im Folgenden: ds-v3) und deepseek-reasoner (deepseek-r1-0120, im Folgenden: ds-r1)[11]. Für jedes Experiment wurden dem Modell jeweils zehn Beispielsätze als Few-Shot-Beispiele über die Rolle „Assistant“ beim API-Aufruf bereitgestellt.

Quantitative Ergebnisse

Tabelle 1 auf Seite 14 zeigt die F1-Werte[12] aller Modelle unter Verwendung sämtlicher Prompts für alle sechs Ausdrücke. Tabelle 2 und 3 auf Seite 14 präsentieren jeweils die durchschnittlichen Werte für Precision, Recall und F1-Wert aller Termini sowie Modelle mit verschiedenen Prompts. Insgesamt erzielt die Kombination von Stepwise Frame Prompt, gpt-4.1 und *Regime* unter allen Experimenten den besten F1-Wert.

Effizienz von Prompts

Es zeigt sich generell, dass mit zunehmender Prompt-Komplexität der Recall abnimmt, während die Precision steigt. Der Sense Definition Prompt erwies sich dabei als der beste Performer hinsichtlich des durchschnittlichen F1-Werts. Die Aufgabenbeschreibung des Baseline Prompts, die darauf abzielt zu bestimmen, ob das Wort IR-relevant ist, ist nicht hinreichend differenzierend und führt zu Verwirrung, sodass das Modell unzureichend zwischen verschiedenen semantischen Ebenen unterscheiden kann. Dies hat zur niedrigen Leistung des Begriffs *Regime* mit dem Baseline-Prompt geführt, da sich *Regime* im Sinne einer nationalen Regierung ebenfalls auf die IB beziehen könnte. Ein weiterführendes Experiment, bei dem die Aufgabe dahingehend modifiziert wurde zu bestimmen, ob das Wort ein IR-Fachbegriff ist, erzielte mit ds-v3 – das die niedrigste Leistung in diesem Prompt-Setting aufwies – einen F1-Wert von 98,3 %. Beim Stepwise Frame Prompt hingegen treten – entgegen der Intuition – mehr Fehler auf. Das liegt daran, dass die Modelle durch die Bereitstellung zusätzlicher Informationen und das zweistufige Selektionsdesign mehr Möglichkeiten haben, Fehler zu machen. Wenn ein Modell bei der Identifikation der Frames im ersten Schritt oder bei der Erkennung von *fillers* im zweiten Schritt fehlerhafte Ergebnisse liefert, wirkt sich das unmittelbar auf das endgültige Klassifikationsergebnis aus.

Modellleistung

Generell lässt sich beobachten, dass Reasoning-Modelle keine Überlegenheit gegenüber Chat-Modellen zeigen und das Modell gpt-4.1 die beste Performance aufweist, gefolgt von dem Modell ds-v3. Obwohl das Modell gpt-o4-mini eine höhere durchschnittliche Precision gegenüber dem Modell gpt-4.1 hat, ist es in anderen Evaluationsmetriken deutlich unterlegen.

Das Modell gpt-4.1 erzielt den besten F1-Wert mit dem Stepwise Frame Prompt und zeigt somit eine mögliche Überlegenheit bei der Identifikation von *fillers*. Dies ist jedoch durch weitere Experimente und Fehleranalysen zu bestätigen.

Verschiedene Termini

Die unterschiedlichen Prompt-Modell-Kombinationen zeigen

[9] Da im vorliegenden Beitrag die Gestaltung von Frames das primäre Anliegen darstellt, werden auf der Basis der im FrameNet (vgl. NLTK FrameNet-Korpus) vordefinierten Frames geringfügige Modifikationen vorgenommen, wobei besondere Berücksichtigung den zentralen Frame-Elementen für die Klassifikation gilt.

[10] Vgl. https://platform.openai.com/docs/models.

[11] Vgl. https://api-docs.deepseek.com/quick_start/pricing.

[12] F1-Wert, Precision und Recall sind Metriken zur Bewertung von Klassifikationsergebnissen. „Recall“ bezieht sich darauf, wie viele der tatsächlich positiven Instanzen korrekt als positiv klassifiziert wurden, und „Precision“ drückt aus, wie viele der als positiv klassifizierten Instanzen tatsächlich positiv sind. Der F1-Wert stellt das harmonische Mittel von Precision und Recall dar.

Deutscher
Terminologie-Tag e.V.

Deutsches
Institut für Terminologie e.V.

40 Jahre DTT

Feiern Sie mit uns auf dem

20. DTT-Symposion

vom 17.–19. Juni 2027

Termin jetzt schon vormerken!

termspezifische Variationen in ihren Ergebnissen. Das erschwert eine Generalisierung dieser Methode. Tabelle 2 zeigt, dass der Sense Definition Prompt für die Termini *Entspannung*, *Intervention* und *Norm* bessere Ergebnisse erzielt. Bei den Begriffen *Kooperation* und *Regime* ist hingegen der Stepwise Frame Prompt überlegen, bei *Integration* der Baseline Prompt.

Es lässt sich auch beobachten, dass die durchschnittlichen F1-Werte von Termini wie *Integration* und *Norm* generell niedriger ausfallen als die anderen. Offenbar begegnen die Modelle bei der Verarbeitung dieser Ausdrücke ähnlichen Schwierigkeiten wie menschliche Annotierende.

So trägt beispielsweise *Integration* im IB-Kontext die Bedeutung der Entstehung neuer internationaler Akteure in sich. Folglich erweist sich die Unterscheidung zwischen der Integration in die Weltwirtschaft und dem Beitritt zu bereits bestehenden internationalen Organisationen als besonders herausfordernd.

Reflexion und Ausblick

Wenn wir die Ergebnisse mit Sense Definition und Stepwise Frame Prompt betrachten, zeigt der hier dargestellte Ansatz sein Potenzial. Im Vergleich zu anderen etablierten Methoden erzielen die Klassifikationsexperimente bereits hervorragende Ergebnisse bei geringem Aufwand. Die inhärente Komplexität der IB-Terminologie macht es nahezu unmöglich, bei dieser Aufgabe hundertprozentige Ergebnisse zu erzielen. Auch menschliche Annotierende geraten in manchen Zweifelsfällen an ihre Grenzen. Da die Interpretation und die Frame-semantische Modellierung aller Ausdrücke manuell erfolgen und eine vollständige Vorhersage aller spezifischen Anwendungsfälle nicht möglich ist, können Fehler auch dadurch entstehen. Eine präzisere Beschreibung der Bedeutungen sowie die Auswahl weiterer repräsentativer Beispielsätze könnte die Modellleistung weiter verbessern. Darüber hinaus treten bei generativen LLMs Konsistenzprobleme auf, d. h., die Ausgaben des Modells können bei wiederholten Durchläufen variieren. Daher empfiehlt sich die mehrfache Ausführung, um die Ergebnisse zu validieren. Zusätzlich wäre ein Finetuning des Modells mittels annotierter Beispielsätze in Betracht zu ziehen.

Es muss noch eingeräumt werden, dass die auf binäre Klassifikation zielenden Beschreibungen in den Prompts nicht vollständig die Komplexität der IB-Terminologie erfassen, wie wir sie oben dargestellt haben. Angesichts des Ziels, sämtliche polysemen Wörter zu identifizieren, die mit hoher Wahrscheinlichkeit als Fachbegriffe verwendet werden, um weitere Analysen zu ermöglichen, erscheint eine solche grobkörnige Filterung aber angemessen.

Für weiterführende Untersuchungen – wie die detaillierte Analyse der Klassifikationsfehler, die Reflexion des Reasoning-Prozesses sowie die Prüfung, inwieweit diese Methode die spezifischen Kontexte identifizieren kann, in denen ein Wort als Fachbegriff verwendet wird (d. h. die Füllung der Frame-Elemente) – bedarf es jedoch zukünftiger tiefergehender Forschung.

Literaturverzeichnis

[1] Basile, Pierpaolo; Siciliani, Lucia; Musacchio, Elio; Semeraro, Giovanni (2025): „Exploring the Word Sense Disambiguation Capabilities of Large Language Models“. URL: https://arxiv.org/pdf/2503.08662 (Zugriff: 09.07.2025).

[2] Bellers, Jürgen (1999): „Integration“. In: Boeckh, Andreas (Hrsg.): Lexikon der Politik. Internationale Beziehungen. Frankfurt a. M.: Büchergilde Gutenberg, S. 216-220.

[3] Bevilacqua, Michele; Pasini, Tommaso; Raganato, Alessandro; Navigli, Roberto (2021): „Recent Trends in Word Sense Disambiguation. A Survey“. In: IJCAI-21: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence. S. 4330-4338.

[4] Busse, Dietrich (2017): Lexik. Frame-analytisch. In: Niehr, Thomas; Kilian, Jörg; Wengeler, Martin (Hrsg.): Handbuch Sprache und Politik, Bd. 1. Bremen: Hempen Verlag, S. 194-220.

[5] Fillmore, Charles J. (1982): Frame Semantics. In: Linguistic Society of Korea (Hrsg.): Linguistics in the Morning Calm. Seoul: Hanshin Publishing Co., S. 111-137.

[6] Jahn, Egbert (1999): „Integration“. In: Boeckh, Andreas (Hrsg.): Lexikon der Politik. Internationale Beziehungen. Frankfurt a. M.: Büchergilde Gutenberg, S. 259-260.

[7] Kang, Haoqiang; Blevins, Terra; Zettlemoyer, Luke (2025): „Translate to Disambiguate. Zero-shot Multilingual Word Sense Disambiguation with Pretrained Language Models“. URL: https://arxiv.org/pdf/2304.13803 (Zugriff: 09.07.2025).

[8] Krasner, Stephan (1982): „Structural Causes and Regime Consequences. Regimes as Intervening Variables“. In: International Organization. Nr. 36 (2), S. 185-205.

[9] Liu, Pengfei; Yuan, Weizhe; Fu, Jinlan; Jiang, Zhengbao; Hayashi, Hiroaki; Neubig, Graham (2023): „Pretrain, Prompt, and Predict. A Systematic Survey of Prompting Methods in Natural Language Processing“. In: ACM Computing Surveys. Nr. 55(9), S. 1-35.

[10] Loureiro, Daniel; Rezaee, Kiamehr; Pilehvar, Mohammad T.; Camacho-Collados, Jose (2021): „Analysis and Evaluation of Language Models for Word Sense Disambiguation“. In: Computational Linguistics. Nr. 47(2), S. 387-443.

[11] Motos, Raquel M. (2013): „The Role of Interdisciplinarity in Lexicography and Lexicology“. In: Balteiro, Isabel (Hrsg.): New Approaches to Specialised English Lexicology and Lexicography. Newcastle: Cambridge Scholars Publishing, S. 3-13.

[12] Müller, Marcus; Steffek, Jens; Schenk, Anna (2020): „The Darmstadt International Relations Corpus (DIReC)“. URL: https://tuprints.ulb.tu-darmstadt.de/13063/1/Direc_working%20paper_20200730.pdf (Zugriff: 21.06.2025).

[13] Müller, Marcus; Mell, Ruth M. (2020): „Zwischen Fach und Wort. Fragen, Methoden und Erkenntnisse der Terminologiedynamik“. In: Bopp, Dominika; Pthashnik, Stefanyia; Roth, Kerstin; Theobald, Tina (Hrsg.): Wörter. Zeichen der Veränderung. Berlin, Boston: De Gruyter, S. 191-208.

[14] Mukherjee, S. P.; Sinha, Bikas K.; Chattopadhyay, Asis K. (2018): Statistical Methods in Social Science Research. Singapore: Springer.

[15] NIC (2021): „US-Backed International Norms Increasingly Contested“. URL: https://www.dni.gov/files/images/globalTrends/GT2040/NIC-2021-02491_GT_Future_of_Int_Norms_22Mar22_UNSOURCED.pdf (Zugriff: 01.04.2025).

[16] Nohlen, Dieter; Schultze, Rainer-Olaf; Schüttemeyer, Suzanne S. (Hrsg.) (1999): Lexikon der Politik. Politische Begriffe. Frankfurt a. M.: Büchergilde Gutenberg.
[17] Onuf, Nicholas (1989): World of Our Making. Rules and Rule in Social Theory and International Relations. Columbia: S.C.
[18] Otto, Carsten (1997): Das Waffenregister der Vereinten Nationen. Wiesbaden: Springer.
[19] Schubert, Klaus; Klein, Martina (2020): „Entspannungspolitik". In: Das Politiklexikon. Bonn: Dietz. Lizenzausgabe Bonn: Bundeszentrale für politische Bildung. URL: https://www.bpb.de/kurz-knapp/lexika/politiklexikon/17407/entspannungspolitik/ (Zugriff: 17.05.2025).
[20] Schnedermann, Theresa (2021): Die Macht des Definierens. Eine diskurslinguistische Typologie am Beispiel des Burnout-Phänomens. Berlin, Boston: De Gruyter.
[21] Steffek, Jens; Müller, Marcus; Behr, Hartmut (2021): „Terminological Entrepreneurs and Discursive Shifts in International Relation. How a Discipline Invented the ‚International Regime'". In: International Studies Review. Nr. 23(1), S. 30-58.
[22] Sumanathilaka, T.G.D.K.; Micallef, Nicholas; Hough, Julian (2024): „Can LLMs assist with Ambiguity? A Quantitative Evaluation of various Large Language Models on Word Sense Disambiguation". In: Proceedings of the 1st International Conference on NLP & AI for Cyber Security. S. 97-108.
[23] Temmerman, Rita (2000): Towards New Ways of Terminology Description. The Sociocognitive Approach. Amsterdam, Philadelphia: John Benjamins.
[24] Vergerio, Claire (2019): „Context, Reception, and the Study of Great Thinkers in International Relations". In: International Theory. Nr. 11(1), S. 110-137.
[25] Wendt, Alexander (1999): Social Theory of International Politics. Cambridge: Cambridge University Press.
[26] Yae, Jung H.; Skelly, Nolan C.; Ranly, Neil C.; LaCasse, Phillip M. (2025): „Leveraging Large Language Models for Word Sense Disambiguation". In: Neural Computing and Applications. Nr. 37, S. 4093-4110.

Mo Han promoviert an der Technischen Universität Darmstadt und arbeitet als wissenschaftliche Mitarbeiterin für das Projekt „Terminologische Innova-tionen in Internationale Beziehungen". Ihr Fokus liegt auf Terminologiearbeit und Korpuslinguistik.

Kontaktadresse
mo.han@tu-darmstadt.de
www.linglit.tu-darmstadt.de/institutlinglit/mitarbeitende/han__mo/standardseite_han_linglit.de.jsp

Marcus Müller ist Professor für Germanistik - Digitale Linguistik am Institut für Linguistik und Literaturwissenschaft der TU Darmstadt. Seine Forschung liegt an der Schnittstelle von Korpuslinguistik, Diskursanalyse und Lexikogrammatik. Ein Schwerpunkt dabei ist die empirische Terminologieforschung in Wissenschaftsdiskursen. Marcus Müller leitet das Discourse Lab an der TU Darmstadt, eine kollaborative Forschungs- und Lehrplattform für digitale Diskursanalyse und Korpuslinguistik.

Kontaktadresse
marcus.mueller@tu-darmstadt.de
www.linglit.tu-darmstadt.de/institutlinglit/mitarbeitende/marcusmueller/

Semantische Suche und erklärbare KI – Neue Wege in der Literaturrecherche

Sinan Sen

Keyword-based methods continue to dominate document search, including in academic research, while semantic approaches remain underused. We present a search engine that interprets free-text queries and identifies relevant publications based on their semantics. Designed for efficiency and adaptability, it incorporates Explainable Artificial Intelligence (XAI) to improve literature retrieval. This not only enhances transparency and accelerates evaluation but also provides a new perspective on discovering academic research documents.

Keywords: Semantic Search, Literature Research, Explainable AI, BERT, Academic Search Engine

Wissensarbeiter verarbeiten Informationen, schaffen neues Wissen und lösen komplexe Probleme. Für Unternehmen, Universitäten und öffentliche Einrichtungen sind sie heute unverzichtbar, denn sie treiben Fortschritte in Wissenschaft, Technik und Gesellschaft voran. Entsprechend gewinnt die gezielte Förderung ihrer Produktivität zunehmend an Bedeutung. Technische Verfahren der Textanalyse und Künstlichen Intelligenz (KI) haben das Potenzial, Wissensarbeiter spürbar zu entlasten, indem sie Routinetätigkeiten reduzieren, Informationen schneller zugänglich machen und so den kreativen und analytischen Anteil der Wissensarbeit stärken.

Schon 2001 schätzte die International Data Corporation (IDC), dass Wissensarbeiter rund 30 % ihrer Arbeitszeit mit der Suche nach Informationen verbringen [1]. Spätere Studien zeigten zudem, dass Recherchen ohne Vorwissen deutlich länger dauern und ein erheblicher Teil der Zeit damit verloren geht, Dokumente zu sichten und zu bewerten [2,3]. Informationen sind also vorhanden, jedoch ist der Zugang dazu oft ineffizient.

Ein zentrales Problem klassischer Suchverfahren ist ihre syntaxorientierte Arbeitsweise: Treffer entstehen nur, wenn die gesuchten Begriffe exakt oder in sehr ähnlicher Form im Dokument vorkommen. Synonyme oder inhaltlich verwandte Begriffe bleiben unberücksichtigt. Zwar helfen boolesche Suchbefehle in gewissem Umfang, doch auch sie setzen voraus, dass man die richtigen Begriffe kennt.

In unserer Arbeit haben wir dieses Problem am Beispiel der Literaturrecherche untersucht. Dazu haben wir eine akademische Suchmaschine entwickelt, die auf einem eigens trainierten BERT-Sprachmodell basiert (BERT – ein von Google entwickeltes KI-Modell, das Texte nicht nur wortwörtlich, sondern auch ihrem Sinn nach versteht). Mit dieser Suchmaschine zeigen wir, wie semantische Suche neue Möglichkeiten eröffnet:

- **Überblick gewinnen:** Das System erkennt Schlüsselbegriffe in Publikationen und zeigt Themenfelder, die Forschende bereits als relevant betrachten. Dies bietet einen strukturierten Einstieg in neue Wissensgebiete.
- **Inhalte erschließen:** Freitextanfragen werden semantisch mit Abstracts und Titeln abgeglichen. Dadurch schlägt das System Publikationen vor, die nicht nur syntaktisch, sondern auch inhaltlich passen.

Ergänzt wird unser Ansatz durch Methoden der erklärbaren Künstlichen Intelligenz (Explainable AI, XAI). Sie machen die Auswahl der Treffer nachvollziehbar, erleichtern die Bewertung der Ergebnisse und unterstützen den gesamten Rechercheprozess.

Für Wissensarbeiter ergeben sich daraus neue Möglichkeiten: Semantische Ansätze erweitern klassische Suchmethoden, insbesondere wenn sie durch domänenspezifisch trainierte Modelle gestützt werden. Die Literaturrecherche ist ein exemplarisches Anwendungsfeld, der Ansatz lässt sich jedoch auch auf andere Dokumentarten übertragen.

Semantische Suche mit Transformer-Modellen

Transformer-Modelle wurden 2017 mit der Arbeit *Attention Is All You Need* von Vaswani et al. eingeführt [4]. Ihr zentrales Prinzip ist der Self-Attention-Mechanismus: Jedes Wort wird nicht isoliert, sondern im Kontext aller anderen gleichzeitig betrachtet. Damit unterscheiden sich

Transformer deutlich von älteren Verfahren, die Texte sequenziell Wort für Wort verarbeiteten und dadurch langsamer sowie weniger geeignet für längere Passagen waren.

Seitdem haben Transformer die Sprachverarbeitung (NLP – Natural Language Processing) stark geprägt, von maschineller Übersetzung über Sentimentanalyse bis hin zur automatischen Beantwortung von Fragen [5]. Besonders prägend sind Modelle wie BERT, das auf Sprachverständnis spezialisiert ist [6], und GPT, das vor allem für die Textgenerierung eingesetzt wird [7]. Beide basieren auf der ursprünglichen Architektur von 2017 und bilden das Fundament moderner Sprachmodelle (Large Language Models, LLMs) [8].

Heute finden Transformer-Modelle auch in der Literaturrecherche Anwendung, etwa in Suchsystemen wie *Consensus* oder *Elicit*. Diese verbinden klassische Schlüsselwortsuche mit semantischen Verfahren und schlagen dadurch relevantere Ergebnisse vor. Ein offenes Problem bleibt jedoch die mangelnde Erklärbarkeit: Nutzer erfahren nur selten, warum bestimmte Publikationen vorgeschlagen werden. Forschungsarbeiten aus dem Bereich Explainable AI setzen hier an und versuchen, mehr Transparenz zu schaffen [9].

Von Rechercheproblemen zu Systemanforderungen

Um Anforderungen an eine Suchmaschine zu ermitteln, führten wir leitfadengestützte Interviews mit sechs wissenschaftlichen Mitarbeitenden aus unterschiedlichen Disziplinen und Karrierestufen. Keiner der Teilnehmenden arbeitete im KI-Bereich, und alle verfügten nur über oberflächliche Kenntnisse zu diesem Thema.

Die Gespräche behandelten typische Vorgehensweisen bei der Literaturrecherche, alternative Strategien, wiederkehrende Schwierigkeiten sowie die Vorstellung eines idealen Unterstützungssystems.

Dabei zeigte sich: Viele Recherchen verlaufen eher unstrukturiert, Strategien werden flexibel angepasst, aber selten bewusst auf das jeweilige Szenario zugeschnitten. Häufig beschränken sich erste Sichtungen auf Titel und Abstracts, was die Entscheidung über die Relevanz erschwert. Besonders problematisch ist die Recherche in interdisziplinären Feldern oder in Themengebieten mit wenig Vorwissen. Zudem äußerten die Befragten den Wunsch nach zusätzlichen Kontextinformationen zu Treffern, um schneller einschätzen zu können, ob ein Artikel relevant ist.

Aus diesen Beobachtungen leiteten wir drei zentrale Anforderungen an das zu entwickelnde Suchsystem ab:

1. **Natürliche Spracheingabe:** Nutzer sollen Freitextanfragen formulieren können, die semantisch verarbeitet werden, ohne die exakten Fachterminologien syntaktisch zu kennen oder boolesche Operatoren zu verwenden.
2. **Begriffsempfehlungen:** Das System soll verwandte oder fachübergreifende Begriffe vorschlagen, die die Recherche erweitern.
3. **Erklärbarkeit:** Das System soll transparent machen, warum Ergebnisse der Freitextanfragen als relevant eingestuft wurden, um die Nachvollziehbarkeit zu erhöhen.

Semantische Suche durch Vektorrepräsentation

Um die erste zentrale Anforderung, die Ablösung einer rein schlüsselwortbasierten durch eine semantische Suche, zu erfüllen, haben wir ein eigens trainiertes Sprachmodell entwickelt. Es basiert auf einem Transformer, der Texte in sogenannte Embeddings überführt, also in Vektoren, die den Sinn eines Textes mathematisch abbilden [4]. Der Vorteil dieser Repräsentation besteht darin, dass inhaltlich ähnliche Texte im selben Bereich des Vektorraums liegen, auch wenn unterschiedliche Begriffe verwendet werden.

Beispiel

In einem semantischen Vektorraum liegen Begriffe nicht isoliert, sondern entsprechend ihrer inhaltlichen Nähe zueinander. So befindet sich „artificial neural networks“ in unmittelbarer Nähe zu Begriffen wie „deep learning“ oder „classification algorithms“. Obwohl diese Ausdrücke unterschiedlich formuliert sind, ordnet das Modell sie im selben Bereich an, weil sie thematisch stark verbunden sind. Klassische Verfahren würden diese Zusammenhänge nicht erkennen und die Begriffe unabhängig voneinander behandeln.

Die Vektorrepräsentation bildet damit die Grundlage für eine semantische Suche. Um jedoch ein Modell zu entwickeln, das speziell für wissenschaftliche Publikationen geeignet ist, gehen wir im nächsten Schritt auf die Details unserer Implementierung mit LLMs ein.

Domänenspezifische Sprachmodelle für die semantische Suche in Publikationen

Die Vorgehensweise beim Training von LLMs besteht in der Regel aus zwei Schritten: (1) einem allgemeinen Pre-Training und (2) einem domänenspezifischen Fine-Tuning.

Im ersten Schritt, dem Pre-Training, werden die Modelle auf sehr großen allgemeinen Textsammlungen wie Wikipedia, Büchern oder Webtexten trainiert. Dabei lernen sie grundlegende Sprachmuster, Grammatik und semantische Strukturen [4]. Dieses breite Wissen macht Modelle wie BERT oder GPT zu universellen Sprachsystemen, die in vielen Kontexten einsetzbar sind. In diversen Anwendungsfällen können solche vortrainierten Modelle bereits

ausreichen, etwa wenn nur ein allgemeines Sprachverständnis erforderlich ist.

Für eine präzise semantische Suche würden diese jedoch nicht genügen, da die Modelle nicht auf die Sprache und Schwerpunkte spezifischer Publikationspools zugeschnitten sind. Auch im Hinblick auf die Erweiterung um erklärbare KI ist eine domänenspezifische Anpassung notwendig, damit Relevanzbewertungen und Einflussfaktoren zuverlässig nachvollziehbar gemacht werden können.

Daher folgt als zweiter Schritt das Fine-Tuning auf unseren eigenen, kuratierten Sammlungen aus IEEE, Springer und Scopus. So entsteht ein Modell, das die Terminologie, Abkürzungen und Ausdrucksweisen dieser Texte gezielt erlernt. Der zentrale Unterschied liegt somit in der Fokussierung und Anpassung: Während das Pre-Training ein breites Sprachverständnis vermittelt, richtet das Fine-Tuning das Modell passgenau auf die für unsere Recherche relevanten Daten aus [10].

Für das Fine-Tuning haben wir 350.000 Publikationen aus IEEE, Springer und Scopus gesammelt, jeweils bestehend aus Titel, Abstract und Schlagwörtern. Um das Fine-Tuning handhabbar zu gestalten, nutzten wir zunächst 150.000 Einträge zum Thema Künstliche Intelligenz. Ziel war es nicht, ein universelles Sprachmodell nachzubauen, sondern ein auf wissenschaftliche Texte spezialisiertes Modell zu entwickeln, das die Sprache von Abstracts und Fachartikeln präziser versteht und im Rahmen eines Demonstrators sichtbar gemacht wird.

Als Ausgangsbasis für das Fine-Tuning wählten wir DistilBERT, eine ressourcenschonendere Variante des bekannten BERT-Modells [11]. DistilBERT bringt bereits durch sein Pre-Training ein breites Sprachverständnis mit. Dieses Grundwissen bildet die Basis, auf der wir mit unserem spezifischen Publikationspool weiterarbeiten.

Für die Anpassung nutzten wir Contrastive Learning [12,13]. Dieses Verfahren eignet sich, um semantische Ähnlichkeiten zwischen Texten zu erfassen, ohne manuell annotierte Daten zu benötigen. Dabei werden Paare von Texten gebildet:

- **Positive Paare** bestehen aus Titel und Abstract derselben Publikation.
- **Negative Paare** bestehen aus Titel und Abstract unterschiedlicher Publikationen.

Das Modell lernt, die Embeddings der positiven Paare enger zusammenzuführen und die der negativen Paare auseinanderzuhalten.

Beispiel

Steht im Titel „Convolutional Neural Networks“ und das zugehörige Abstract enthält Begriffe wie „image classification“ oder „computer vision“, erkennt das Modell die inhaltliche Nähe. Wird derselbe Titel jedoch mit einem Abstract über „distributed databases“ oder „network security“ kombiniert, stuft das Modell diese Kombination als unpassend ein.

Während das Pre-Training ein breites Verständnis von Sprache und Bedeutungen vermittelt, richtet das Fine-Tuning das Modell auf die Sprache wissenschaftlicher Abstracts aus. Contrastive Learning schärft diesen Prozess, indem es inhaltlich zusammengehörige Texte näher zusammenführt und unpassende klar trennt. Auf diese Weise erkennt die Suchmaschine nicht nur identische Wörter, sondern auch inhaltlich verwandte Publikationen.

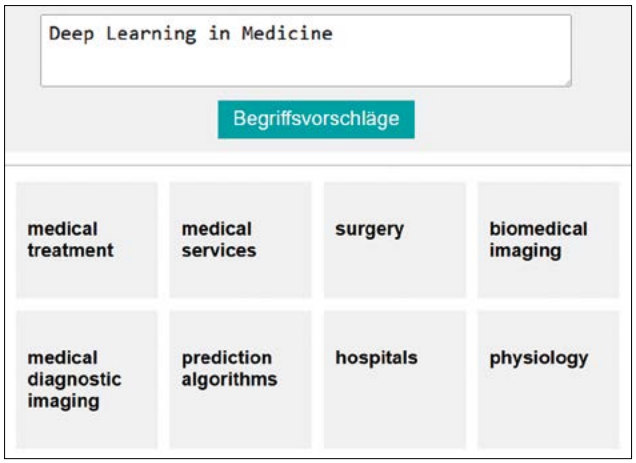


Abb. 1: Die Abbildung illustriert das Verfahren der Term-Empfehlungen: Ausgehend von der Eingabe „Deep Learning in Medicine“ werden semantisch ähnliche Begriffe wie „medical treatment“, „medical diagnostic imaging“ oder „biomedical imaging“ vorgeschlagen. (Quelle: eigener Screenshot)

Begriffsempfehlungen

Zur Umsetzung der zweiten Systemanforderung wurde eine Funktion zur Empfehlung verwandter Begriffe integriert. Dafür haben wir die Keywords aller Publikationen in unserer Datenbank zusammengeführt und mithilfe derselben Embedding-Technologie semantisch repräsentiert. Diese Repräsentationen werden mit den Eingaben der Nutzer verglichen, sodass das System inhaltlich nahe Begriffe vorschlagen kann.

Beispiel (siehe Abbildung 1):

- Eingabe: *„Deep Learning in Medicine“*
- Empfohlene Begriffe: *„medical treatment“*, *„medical diagnostic imaging“*, *„biomedical imaging“*

Auf diese Weise werden Nutzer auf relevante Fachtermini hingewiesen, die ihnen möglicherweise nicht bekannt sind, die aber im jeweiligen Forschungsfeld eine zentrale Rolle spielen. Dies erleichtert besonders interdisziplinäre Recherchen und erweitert den Zugang zu neuem Wissen.

6. März bis 20. Juni 2026
Onlinekurs mit Hybridveranstaltungen am SDI in München

SDI INTERNATIONALE HOCHSCHULE
UNIVERSITY OF APPLIED SCIENCES

KÜNSTLICHE INTELLIGENZ IN DER ERBRINGUNG VON ÜBERSETZUNGSDIENSTLEISTUNGEN

Ein Zertifikatskurs des BDÜ in Kooperation mit dem SDI München

Onlinekurs für Übersetzer/-innen, die die nötigen Kenntnisse erwerben möchten, um KI-Systeme zur Verarbeitung natürlicher Sprache effizient und verantwortungsvoll zu nutzen und KI-Anwendungen für den Einsatz in Übersetzungsprozessen zu konzipieren und zu implementieren.

In **knapp 100 Unterrichtsstunden** vermittelt der Kurs

- theoretische Grundlagen
- Kenntnisse (einschließlich Grundkenntnisse der Programmierung in Python) und Fertigkeiten
  - zum Einsatz von KI im Terminologiemanagement
  - zum Aufbau von Textkorpora für Large Language Models
  - zu Pre- und Post-Editing

Weitere Themen sind Qualitätskontrolle, Fehlermessverfahren und rechtliche Aspekte der KI-Nutzung. Die Unterrichtseinheiten finden im Zeitraum vom **06.03. bis 20.06.2026** freitagnachmittags und samstagvormittags **online** statt.

Die Kursblöcke am **06./07.03.2026** und am **19./20.06.2026** finden am SDI in München als **Hybridveranstaltungen** statt, an denen die Kursteilnehmer/-innen **wahlweise in Präsenz oder online** teilnehmen können.

Weitere Infos zu Inhalt, Referent/-innen und Anmeldung unter: seminare.bdue.de/6721

NEUERSCHEINUNG

Translation Management mit KI und LangOps:
Von Augmented Translation und Postediting bis zu Prozessmodellierung und Language Orchestration

In Zusammenarbeit mit der Internationalen Hochschule SDI München hat **Prof. Dr. Rachel Herwartz (Hrsg.)** in diesem Sammelband die Fachbeiträge von insgesamt zehn Autorinnen und Autoren zum Themenbereich Translation Management mit KI und LangOps veröffentlicht.

- **Bestellung:** www.bdue-fachverlag.de/detail_book/194
- **Leseprobe:** www.bdue-fachverlag.de/download/books/5986

ISBN: 978-3-946702-46-7, 257 Seiten, Erscheinungsjahr: 2025, Preis: 49,00 €

Weiterbildung für Dolmetscher und Übersetzer

Immer auf dem neuesten Stand – technisch, fachlich, sprachlich, unternehmerisch:
www.bdue-fachverlag.de

BDÜ

Erklärbarkeit der Semantischen Suche

Ein zentrales Problem moderner KI-Systeme ist ihre Intransparenz. Sie liefern zwar Ergebnisse, doch bleibt für Nutzer oft unklar, wie diese zustande kommen. Gerade im wissenschaftlichen Umfeld ist das kritisch, weil Vertrauen, Nachvollziehbarkeit und die aktive Auseinandersetzung mit Inhalten entscheidend für Wissensbildung sind. Dieses Grundproblem beschreiben Doshi-Velez und Kim als zentrale Herausforderung erklärbarer KI [14].

Um dieses Problem zu adressieren, haben wir die semantische Suchfunktionalität um eine Erklärbarkeits-Funktion erweitert. Sie macht sichtbar, welche Begriffe einer Anfrage den größten Einfluss auf das Ergebnis hatten. Besonders wichtige Wörter werden farbig markiert – stark einflussreiche rot, weniger relevante blau. Ribeiro et al. stellten mit *LIME* („Local Interpretable Model-Agnostic Explanations“) eine Methode vor, die genau diesen Ansatz verfolgt: sichtbar zu machen, welche Eingaben die Entscheidung eines Modells geprägt haben [15]. Auf dieser Grundlage haben wir unser Verfahren entwickelt.

Die Berechnung erfolgt mit einem „leave-one-out“-Verfahren: Die Anfrage wird mehrfach durchgespielt, jedes Mal ohne einen bestimmten Bestandteil. Dabei können je nach Einstellung einzelne Wörter („employment“), Funktionswörter ohne Bedeutung („to“) oder ganze Begriffe wie „Artificial Intelligence“ oder „disabled people“ ausgeblendet werden. Verändert sich die Relevanzbewertung deutlich, wird dieser Ausdruck als besonders einflussreich eingestuft und in der Oberfläche rot markiert. Beispiel: Sinkt die Relevanzbewertung von 90 % auf 40 %, wenn der Begriff „employment“ entfernt wird, zeigt die rote Markierung, dass dieser Begriff entscheidend für das Ergebnis war (siehe Abbildung 2).

Aus den laufenden Evaluationen sowie unseren Beobachtungen im Unternehmens- und akademischen Umfeld zeigt sich: Der Nutzen dieses Ansatzes geht über reine

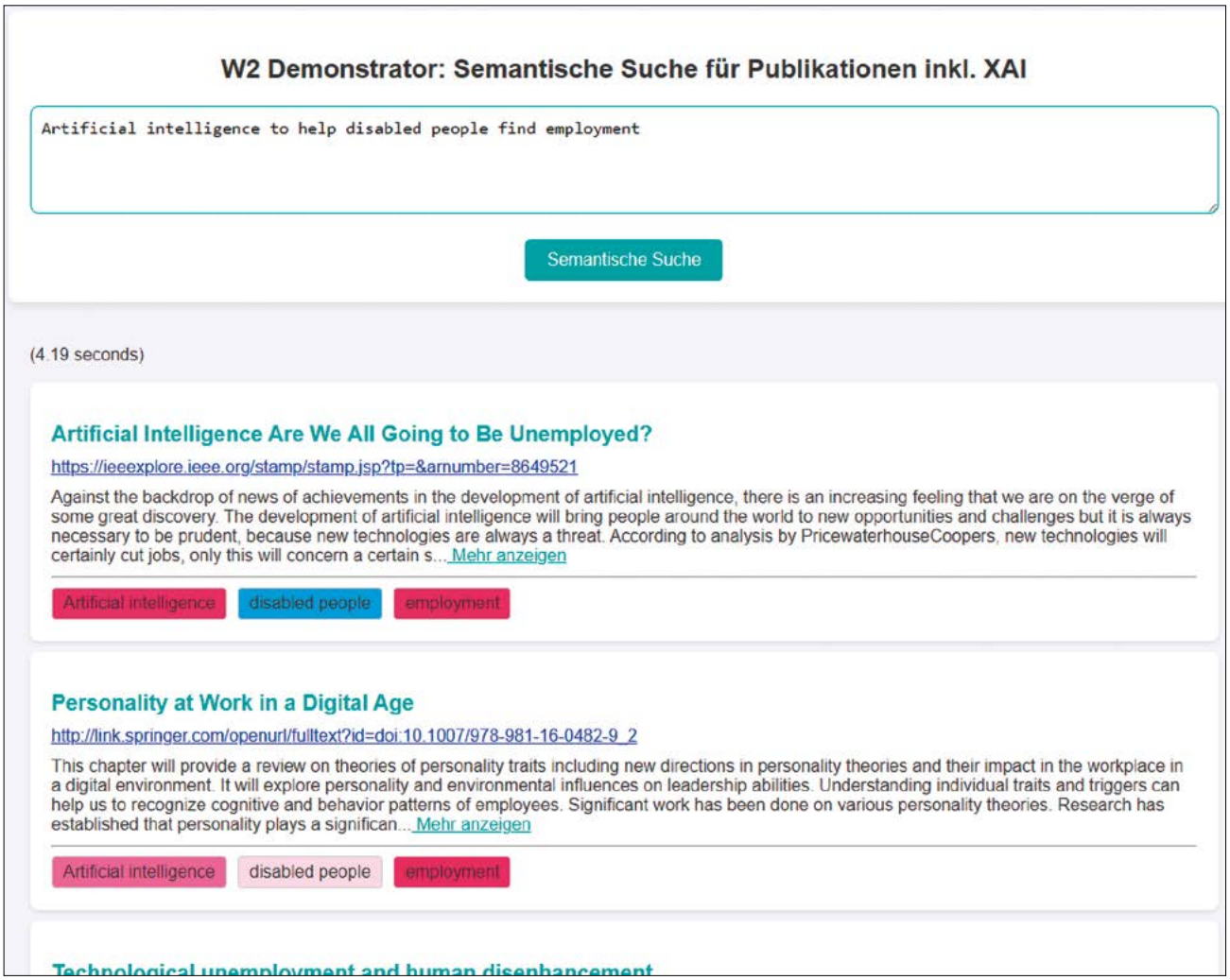


Abb. 2: Eingabe: „Artificial intelligence to help disabled people find employment“. Das Suchergebnis wird durch XAI unterstützt, indem die Begriffe der Anfrage nach ihrer Relevanz markiert werden: stark einflussreiche Terme erscheinen Rot, weniger relevante in Blau. Dadurch wird für die Nutzer sichtbar, welche Bestandteile der Anfrage maßgeblich zur Auswahl des jeweiligen Treffers beigetragen haben. (Quelle: eigener Screenshot)

Transparenz hinaus. Nutzer können ihre Anfragen gezielt anpassen, sich intensiver mit den vorgeschlagenen Inhalten auseinandersetzen und zusätzliche Begriffe in ihr eigenes Wissensnetzwerk integrieren. Auf diese Weise unterstützt Erklärbarkeit nicht nur das Vertrauen in die Suchmaschine, sondern auch aktive Wissensbildung und interdisziplinäre Vernetzung.

Zusammenfassung und Ausblick

Mit unserer Suchmaschine möchten wir zeigen, wie semantische Verfahren in Kombination mit Methoden der Explainable AI neue Möglichkeiten in der Dokumentensuche für Wissensarbeiter eröffnen können. Auch wenn die Implementierung am Beispiel der Publikationssuche erfolgte, liegt ein zentrales Potenzial in der Übertragbarkeit auf weitere Domänen.

Der Demonstrator verarbeitet Freitextanfragen semantisch, empfiehlt thematisch verwandte Begriffe und macht die einflussreichsten Terme einer Anfrage transparent sichtbar. Dadurch entsteht nicht nur mehr Nachvollziehbarkeit, sondern auch die Chance, sich aktiver mit Inhalten auseinanderzusetzen und individuelle Wissensnetzwerke zu erweitern.

Die bisherigen Ergebnisse verdeutlichen, dass solche Verfahren klassische Suchmethoden sinnvoll ergänzen und eine vielversprechende Forschungsrichtung darstellen. Gleichzeitig können sie auf andere Dokumentarten übertragen werden, sofern eine domänenspezifische Anpassung erfolgt. So leisten sie einen Beitrag dazu, den Zugang zu Wissen effizienter, transparenter und nachhaltiger zu gestalten.

Diese Arbeit ist Teil des Projekts *Kompetenzzentrum KARL – Künstliche Intelligenz für Arbeit und Lernen in der Region Karlsruhe*. Das Projekt wird durch das Bundesministerium für Bildung und Forschung (BMBF) im Rahmen des Programms *Zukunft der Wertschöpfung* gefördert (Förderkennzeichen: 02L19C250).

An diesem Teilprojekt waren neben der Datalyxt GmbH drei weitere Institutionen beteiligt:

- Robin Weitemeyer, Hochschule Karlsruhe (HKA)
- Natalie Beyer, Lavrio.solutions GmbH
- Lena Kölmel, Karlsruher Institut für Technologie (KIT)

Literatur

[1] Feldman, S. / Sherman, C. (2001): The high cost of not finding information: An idc white paper. In: KMWorld Magazin, S. 1–10.
[2] Borlund, P. / Deier, S. / Bystroem, K. (2012): What does time spent on searching indicate? In: Proceedings of the 4th information interaction in context symposium, S. 184–193.
[3] Toms, E.G. / Villa, R. / McCay-Peet, L. (2013): How is a search system used in work task completion? In: Journal of Information Science 39(1), S. 15–25.
[4] Vaswani, A. et al. (2017): Attention Is All You Need. NeurIPS.
[5] Young, T. et al. (2018): Recent trends in deep learning based natural language processing. IEEE Access.
[6] Devlin, J. et al. (2019): BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL.
[7] Radford, A. et al. (2019): Language Models are Unsupervised Multitask Learners. OpenAI.
[8] Bommasani, R. et al. (2021): On the Opportunities and Risks of Foundation Models. Stanford CRFM.
[9] Guidotti, R. et al. (2019): A survey of methods for explaining black box models. ACM Computing Surveys.
[10] Howard, J. / Ruder, S. (2018): Universal Language Model Fine-tuning for Text Classification. ACL.
[11] Sanh, V. / Debut, L. / Chaumond, J. / Wolf, T. (2019): DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv:1910.01108.
[12] Hadsell, R. / Chopra, S. / LeCun, Y. (2006): Dimensionality reduction by learning an invariant mapping. In: CVPR. IEEE. (Klassische Einführung in Contrastive Learning / Siamese Networks).
[13] Reimers, N. / Gurevych, I. (2019): Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In: Proceedings of EMNLP-IJCNLP. Association for Computational Linguistics.
[14] Doshi-Velez, F. / Kim, B. (2017): Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
[15] Ribeiro, M. T. / Singh, S. / Guestrin, C. (2016): "Why Should I Trust You?" Explaining the Predictions of Any Classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Dr. Sinan Sen ist Mitgründer und Geschäftsführer der Datalyxt GmbH. Das Unternehmen versorgt Firmen mit präzisen Datenauswertungen aus Dokumenten. Seit 2005 arbeitet er an KI-gestützten Software-Lösungen in Forschung und Praxis, mit Stationen am DFKI, FZI und KIT. Heute entwickelt er Software, die wissenschaftliche Tiefe mit praktischer Entlastung im Arbeitsalltag verbindet.

Kontaktadresse
sinan.sen@datalyxt.com
www.datalyxt.com

Terminologieerzwingung mit generativen LLMs

ChatGPT und Google Gemini als Übersetzungstools

Vanessa Šorak

This article presents a small-scale study investigating the effectiveness of terminology enforcement with ChatGPT and Google Gemini on translations from English into German. The results indicate that common negative side effects known from terminology enforcement in conventional NMT systems, such as grammatical agreement errors, do not occur with generative LLMs. Nevertheless, their terminology enforcement remains unreliable, with both models achieving success rates of approximately 70%.

Keywords: Artificial Intelligence, Generative LLM, Machine Translation, Terminology Enforcement, Glossary

Terminologiemanagement ist ein zentraler Bestandteil von professionellen Übersetzungsprozessen und insbesondere in hochgradig spezialisierten Fachbereichen wie Medizin, Recht oder Technik von hoher Relevanz. In CAT-Tools ist die Verwaltung terminologischer Ressourcen daher eine Standardfunktion. Glossare und Terminologiedatenbanken können jedoch auch bei maschinellen Übersetzungsprozessen zum Einsatz kommen, entweder indem sie bei der Entwicklung spezialisierter Modelle direkt in die Trainingspipelines integriert werden oder indem sie bei generischen Modellen als externe Referenzen hochgeladen werden. In letzterem Fall wird oft von *terminology enforcement* (Terminologieerzwingung)[1] gesprochen, da die hochgeladenen Termini während des Übersetzungsprozesses systematisch „erzwungen“ und probabilistisch determinierte Alternativen dabei ggf. überschrieben werden [1]. Der Hauptvorteil dieser Methode gegenüber der Einbindung von Terminologieressourcen in die Trainingspipeline liegt in ihrer sehr einfachen und schnellen Implementierung, die es Endnutzern u. a. erlaubt, ihre Modelle eigenständig zu spezialisieren und regelmäßig mit neuer Terminologie zu aktualisieren. Die meisten kommerziellen Anbieter von herkömmlichen NMÜ-Systemen, darunter etwa DeepL und Google Translate, bieten diese Funktion daher als zusätzliches Feature an.[2] Dabei ist jedoch zu beachten, dass Terminologieerzwingung bei KI-basierten Modellen eine technisch äußerst komplexe Aufgabe ist, die auch bei modernsten Systemen i.d.R. nicht vollkommen zuverlässig ausgeführt wird. So kommt es immer wieder vor, dass einzelne Glossartermini ohne ersichtlichen Grund überhaupt nicht oder nur teilweise übernommen werden. Zudem kann Terminologieerzwingung auch zu negativen Effekten führen, etwa indem zusätzliche Rechtschreib-, Interpunktions- oder Grammatikfehler generiert werden. Typische Beispiele sind Kleinschreibung eines Glossarterminus am Satzanfang, Einfügung von überflüssigen Leerzeichen vor Interpunktionszeichen sowie grammatikalische Inkongruenz zwischen dem Glossarterminus und dessen Artikel [3,4].[3]

[1] Alternativ: *terminology integration* (Terminologieintegration). Diese Bezeichnung wird jedoch auch für die Integration von Glossaren während der Modellentwicklung verwendet. Um Missverständnissen vorzubeugen, wird hier die zwar etwas weniger verbreitete, dafür aber eindeutigere Bezeichnung *Terminologieerzwingung* verwendet.

[2] Während professionelles Terminologiemanagement in der Humanübersetzung oftmals komplexe Terminologiedatenbanken umfasst, mit denen semantische, taxonomische und ontologische Beziehungen zwischen Begriffen ausgedrückt werden können, sind kommerzielle NMÜ-Systeme meist auf den Upload von nicht-hierarchischen, unidirektionalen Wortpaar-Glossaren in .xlsx oder .csv Format beschränkt [2].

[3] Mehrere Studien legen jedoch nahe, dass es diesbezüglich signifikante Unterschiede zwischen verschiedenen NMÜ-Systemen gibt: So zeigten direkte Vergleiche zwischen DeepL und Google Translate beispielsweise, dass DeepLs Terminologieerzwingung nicht nur wesentlich zuverlässiger ist, sondern auch deutlich weniger Negativeffekte generiert (z. B. [3,4]).

Deutscher
Terminologie-Tag e.V.

Deutsches
Institut für Terminologie e.V.

Terminology Work Best Practices 2.0

Das bewährte Praxishandbuch für Terminologiearbeit und Terminologiemanagement mit dem Know-how zahlreicher Experten aus Industrie und Wissenschaft ist auch **in englischer Sprache** erhältlich.

kompakt | praxisnah | auf den Punkt

Geballtes Terminologiewissen für Unternehmen und Freiberufler

1. Argumentationshilfen
2. Grundsätze und Methoden
3. Benennungen
4. Werkzeuge und Technologien
5. Projekt- und Prozessmanagement
6. Berufsprofile, Anforderungen, Ausbildungsinhalte
7. Urheberrecht an Terminologie
8. Wirtschaftlichkeit

Erhältlich in Deutsch oder Englisch
auf www.dttev.org

Vor diesem Hintergrund stellt sich die Frage, ob generative Large Language Models (LLMs)[4] wie OpenAIs ChatGPT oder Google Gemini, die ersten Studien zufolge eine vergleichbare oder gar bessere Übersetzungsqualität erzielen können [5,6], diesbezüglich einen relevanten Vorteil bieten. Aufgrund ihrer Multifunktionalität ist Terminologieerzwingung in der Regel eine inhärente Funktion dieser Modelle – nach aktuellem Kenntnisstand wurde diese in der übersetzungswissenschaftlichen Forschung bislang jedoch kaum berücksichtigt. Die hier präsentierte Studie befasst sich daher mit der Frage, in welchem Maße generative LLMs Terminologie zuverlässig erzwingen können und ob dabei ähnliche Negativeffekte auftreten wie bei herkömmlichen NMÜ-Systemen.

Zur Methodik

Für die vorliegende Untersuchung wurden zwei englischsprachige Texte der Weltgesundheitsorganisation (WHO) ausgewählt, für die jeweils auch deutsche Übersetzungen verfügbar waren. Aus den beiden Ausgangstexten wurde jeweils ein geeigneter Textabschnitt mit einer Länge von einer bis eineinhalb Seiten ausgewählt und manuell für die Übersetzung aufbereitet. Die ursprüngliche Formatierung der Ausgangstexte wurde dabei größtenteils beibehalten, um zu prüfen, ob diese einen Einfluss auf die Terminologieerzwingung haben könnte. Text #1 befasste sich mit der Behandlung von Stimmstörungen nach einer COVID-19-Erkrankung und umfasste 319 Wörter [7]. Der Text enthielt neben fettgedruckten und farbigen Überschriften auch einige Begriffe in Versalschrift und zwei durch Aufzählungspunkte gegliederte Listen. Text #2 thematisierte die missbräuchliche Behandlung von Frauen bei Geburten in geburtshilflichen Einrichtungen und umfasste 621 Wörter [8]. Der Text enthielt farbige und farbig hinterlegte Schrift.

Die beiden Texte wurden als .docx-Dokumente bei ChatGPT (GPT 5.0) und Google Gemini (2.5 Flash) hochgeladen und mit dem Prompt „Translate this document from English into German.“ ins Deutsche übersetzt. Zu beachten ist dabei, dass sich verschiedene Prompting-Strategien unterschiedlich auf die Übersetzungsqualität auswirken können. Eine Studie von He (2024) kam beispielsweise zu dem Schluss, dass zusätzliche Anweisungen (etwa in Form eines klassischen Übersetzungsauftrages) tendenziell zu schlechteren Ergebnissen führen als der schlichte Befehl, den Text von Sprache A in Sprache B zu übersetzen [9]. Aus diesem Grund wurde für die vorliegende Untersuchung der obenstehende Basis-Prompt verwendet. Die so generierten Vorübersetzungen wurden anschließend mit den deutschen Referenzübersetzungen der WHO verglichen. Geeignete Termini, die sowohl in ChatGPTs als auch in Geminis Übersetzung von der WHO-Referenzübersetzung abwichen, wurden jeweils in ein englisch-deutsches Glossar eingetragen. Die daraus resultierenden Glossare umfassten Adjektive, Verben, Substantive und Mehrwortbenennungen. Dabei wurden bewusst auch allgemeinsprachliche Benennungen aufgenommen, um zu prüfen, ob terminologische Präferenzen auch bei Alternativen mit signifikant höherer Probabilität zuverlässig erzwungen werden, z. B. *respectful – wertschätzend* statt *respektvoll* und *disrespectful – geringschätzig* statt *respektlos*. Glossar #1 enthielt letztlich elf Termini mit insgesamt 14 Okkurrenzen in Ausgangstext #1, darunter beispielsweise *ventilate – intubieren* und *acid reflux – saurer Rückfluss*. Glossar #2 enthielt zwölf Termini mit 26 Okkurrenzen in Text #2. Neben *respectful* und *disrespectful* (s. o.) zählten dazu beispielsweise auch *facility for childbirth – geburtshilfliche Einrichtung* und *maternal health care – Gesundheitsfürsorge für Mütter*.

Diese beiden Glossare wurden anschließend als .xlsx-Datei[5] mit dem Prompt „Retranslate the text using the terminology from this glossary.“ in den jeweils selben Chats hochgeladen, um eine Neuübersetzung unter Terminologieerzwingung zu generieren. Die so erstellten Neuübersetzungen wurden anschließend sowohl auf eine erfolgreiche Integration der Glossartermini als auch auf potenzielle Negativeffekte untersucht. An Text #2 wurde zudem ein zusätzlicher Test durchgeführt, in dessen Rahmen das Ausgangsdokument und das entsprechende Glossar gemeinsam in einem separaten Chat hochgeladen wurden. In diesem Fall erhielten die Modelle den Prompt „Translate this document from English into German using the terminology from the glossary.“ Mit diesem zusätzlichen Test sollte überprüft werden, ob sich die Ergebnisse der Terminologieerzwingung im Vergleich zu der zuvor beschriebenen Zwei-Schritt-Methodik unterscheiden, um potenzielle Verzerrungen durch die Neuübersetzung auszuschließen. Die Ergebnisse dieser Testreihe werden im Folgenden kurz vorgestellt. Das vollständige Untersuchungskorpus sowie die Glossare wurden über das Datenrepositorium Zenodo veröffentlicht [10].

[4] Der Begriff generative LLMs bezeichnet hier multifunktionale Sprachmodelle, die in der Lage sind, eigenständig neue Texte und Inhalte zu generieren. Die kürzlich von DeepL und Google Translate eingeführten LLM-basierten Übersetzungsmodelle sind damit ausdrücklich nicht gemeint.

[5] In einem der Tests mit Google Gemini (Text #2, zweiter Test) musste auf eine .csv-Datei zurückgegriffen werden, da das Modell die zuvor verwendete .xslx-Datei plötzlich nicht mehr akzeptierte.

Terminologieerzwingung mit ChatGPT

Bei ChatGPT ließen sich leichte Unterschiede zwischen Text #1 und Text #2 feststellen. In Text #1 wurden die gewünschten Termini in zehn von 14 Fällen übernommen. Besonders auffällig war dabei, dass die Terminologieerzwingung zu Beginn des Textes durchgehend erfolgreich war und lediglich bei den letzten vier Termini scheiterte. Dies legte zunächst nahe, dass die Effektivität der Terminologieerzwingung gen Textende abnehmen könnte. In Text #2 konnte ein solches Muster jedoch trotz der höheren Textlänge nicht mehr beobachtet werden. Hier wurden die gewünschten Termini in 21 von 26 Fällen übernommen, die Terminologieerzwingung scheiterte also in fünf Fällen. Diese traten im Gegensatz zum ersten Text allerdings an verschiedenen Stellen über den Text verteilt auf. In einem dieser Fälle wurde ein Terminus aufgrund der Auslassung eines Satzteils nicht übernommen, die jedoch bereits in der Vorübersetzung auftrat und keinen relevanten Informationsverlust verursachte. Bei den übrigen vier Fällen blieben die Gründe für das Scheitern unklar. So war die Terminologieerzwingung für den Terminus *respectful – wertschätzend* in vier von fünf Fällen erfolgreich und scheiterte nur an einer einzigen Stelle. Darüber hinaus scheiterte die Terminologieerzwingung für *maternal care, unconsented* und *lack of confidentiality* mit jeweils einer Okkurrenz. Der Zusatztest mit Text #2 fiel interessanterweise etwas negativer aus: Hier wurden die gewünschten Termini lediglich in 16 von 26 Fällen übernommen. Die Terminologieerzwingung scheiterte größtenteils an denselben Stellen wie im vorherigen Test, mit einem bedeutenden Unterschied: Der Terminus *respectful – wertschätzend* wurde diesmal bei keiner der fünf Okkurrenzen erzwungen. Dies könnte darauf hindeuten, dass Termini, für die eine statistisch deutlich wahrscheinlichere Alternativübersetzung existiert, mit einem erhöhten Risiko des Scheiterns verbunden sind. Die oben erwähnte Auslassung eines Satzteils trat in diesem Durchlauf nicht mehr auf, wodurch der darin enthaltene Terminus erfolgreich übernommen werden konnte. Negativeffekte, wie Grammatik- oder Interpunktionsfehler, konnten in keinem der drei Tests beobachtet werden.

Terminologieerzwingung mit Google Gemini

In Geminis Fall sollte zunächst angemerkt werden, dass bereits bei den Vorübersetzungen bedeutende Abweichungen von den Ausgangstexten auffielen. So wurde bei der Übersetzung von Text #1 die Überschrift weggelassen, die Satzreihenfolge an mehreren Stellen verändert, Informationen hinzugefügt bzw. verdoppelt und teilweise auch ausgelassen. Das folgende Beispiel zeigt einen Fall, in dem eine Instruktion verdoppelt und zugleich ein Synonym ausgelassen wurde (Hervorhebung in Fettdruck durch die Autorin):

> Ausgangstext: Gastric reflux (**sometimes called acid reflux or heartburn**) is very common, so avoid eating late at night or eating foods that cause indigestion.
>
> Gemini (Vorübersetzung): **Vermeiden Sie späte Mahlzeiten.** Magenreflux (**auch als Sodbrennen bekannt**) ist sehr verbreitet. Vermeiden Sie es, spät in der Nacht zu essen oder Speisen zu sich zu nehmen, die Verdauungsstörungen verursachen.

Auch in der Übersetzung von Text #2 waren zahlreiche Abweichungen zu beobachten. Diese waren jedoch etwas weniger gravierend und beschränkten sich auf Änderungen der Absatzstruktur – wobei Absätze bisweilen auch mitten im Satz eingefügt wurden – sowie Änderungen der Formatierung durch Fettdruck einzelner Begriffe oder Satzteile. Bei beiden Neuübersetzungen unter Terminologieerzwingung traten dieselben oder sehr ähnliche Abweichungen auf.

In Text #1 wurden die gewünschten Termini in zehn von 14 Fällen übernommen. Die Mehrheit der gescheiterten Fälle betrafen den Terminus *problems with your voice – Stimmstörungen*, der insgesamt drei Okkurrenzen aufwies und in keinem Fall erzwungen wurde. Bei der ersten Okkurrenz war dies auch durch die oben bereits erwähnte Auslassung der Überschrift bedingt. Interessanterweise wurde das deutsche Äquivalent *Stimmstörungen* jedoch an einer anderen Stelle im Text hinzugefügt, obwohl der englische Terminus hier nicht vorkam (Hervorhebung in Fettdruck durch die Autorin):

> Ausgangstext: COVID-19 can cause a sore throat, irritable cough and a sensation of mucus pooling in the throat with the need to clear your throat regularly.
>
> Gemini (Vorübersetzung): Nach COVID-19 können Sie möglicherweise eine schwache, verhauchte oder heisere Stimme haben, insbesondere wenn Sie im Krankenhaus beatmet wurden.
>
> Gemini (Neuübersetzung): Nach COVID-19 können Sie möglicherweise **Stimmstörungen** haben, die sich in einer schwachen, hauchigen oder heiseren Stimme äußern, insbesondere wenn Sie im Krankenhaus intubiert wurden.

Darüber hinaus scheiterte die Terminologieerzwingung für *gastric reflux* bei einer von zwei Okkurrenzen. An anderer Stelle wurde ein Glossareintrag zwar erfolgreich übernommen, nicht jedoch seine ursprüngliche Formatierung

(Versalschrift). Abgesehen davon konnten keine weiteren negativen Effekte durch die Terminologieerzwingung beobachtet werden.

In der Neuübersetzung von Text #2 wurden die gewünschten Termini in 18 von 26 Fällen übernommen. Das Scheitern der Terminologieerzwingung war erneut über den Text verteilt zu beobachten, wobei kein klares Muster zu erkennen war. Auffällig war lediglich, dass der Terminus *respectful – wertschätzend* ebenso wie bei ChatGPTs Zusatztest bei keiner seiner fünf Okkurrenzen erzwungen wurde. Darüber hinaus scheiterte die Terminologieerzwingung für *maternal care* und *verbal abuse* mit jeweils einer Okkurrenz sowie *childbirth in facilities* bei einer von drei Okkurrenzen. Absatzstruktur und Formatierung wurden in der Neuübersetzung erneut leicht abgeändert. Die Absatzstruktur der Neuübersetzung wich immer noch deutlich von der des Ausgangstextes ab, es wurden nun aber keine Absätze mehr in fortlaufenden Sätzen eingefügt. Hervorhebungen durch Fettdruck wurden nun überwiegend auf Glossartermini angewandt, jedoch nicht ausschließlich.

Beim Zusatztest mit Text #2 wurden die gewünschten Termini in 20 von 26 Fällen übernommen. Auffällig war dabei, dass der Terminus *respectful – wertschätzend*, der im vorherigen Text bei keiner der fünf Okkurrenzen übernommen wurde, hier in allen Fällen erfolgreich erzwungen wurde. Dafür wurde der Terminus *disrespectful – geringschätzig* nur in zwei von sechs Fällen übernommen. Darüber hinaus scheiterte die Terminologieerzwingung für *childbirth in facilities* in zwei von drei Okkurrenzen. Auch im Zusatztest wich die Absatzstruktur der Übersetzung wieder stark von der des Ausgangstextes ab; es wurden jedoch keine Absätze mehr in fortlaufende Sätze eingefügt. Im Gegensatz zum vorherigen Test gab es hier zudem keine Hervorhebungen durch Fettdruck. Klassische Negativeffekte, wie Grammatik- oder Interpunktionsfehler, konnten in keinem der drei Tests beobachtet werden.

Fazit

Die oben beschriebene Testserie zeigte, dass Terminologieerzwingung auch bei generativen LLMs unzuverlässig ist. ChatGPT und Google Gemini schnitten dabei mit Erfolgsquoten von insgesamt 71 % bzw. 72 % sehr ähnlich ab. Die Gründe für das Scheitern der Terminologieerzwingung blieben in den meisten Fällen unklar. Auffällig war jedoch, dass beide Modelle etwas häufiger Schwierigkeiten mit den Termini *respectful – wertschätzend* und *disrespectful – geringschätzig* hatten. Eine mögliche Erklärung ist, dass in diesen Fällen die statistische Dominanz der alternativen Äquivalente *respektvoll* und *respektlos* priorisiert wurde. Um diese Hypothese zu verifizieren, bedarf es jedoch noch weiterer und umfassenderer Untersuchungen. Die Ergebnisse der Testserie deuten zudem darauf hin, dass bei generativen LLMs keine klassischen Negativeffekte auftreten, wie sie oftmals bei herkömmlichen NMÜ-Systemen beobachtet werden. Gleichzeitig ist jedoch hervorzuheben, dass die Übersetzungen von Google Gemini sowohl mit als auch ohne Terminologieerzwingung teilweise drastisch von den Ausgangstexten abwichen. Solche Abweichungen können in manchen Bereichen, wie beispielsweise bei juristischen Texten, ein ernstzunehmendes Problem darstellen. Die Formatierung des Ausgangstextes scheint bei der Terminologieerzwingung keine relevante Rolle zu spielen.

Wie oben bereits angedeutet, sollten diese Erkenntnisse jedoch noch in weiteren Studien überprüft werden, da die vorliegende Untersuchung aufgrund des sehr kleinen Testkorpus nur eine begrenzte Aussagekraft hat. Um verlässliche Schlüsse über die Effektivität der Terminologieerzwingung bei generativen LLMs ziehen zu können, bedarf es umfassenderer Untersuchungen, die idealerweise auch verschiedene Sprachkombinationen, Textsorten und Prompting-Strategien umfassen. Aufgrund der kontinuierlichen Weiterentwicklung dieser Modelle ist bei allen Untersuchungen dieser Art zudem von einer zeitlich begrenzten Validität der Ergebnisse auszugehen.

Literatur

[1] Matusov, Evgeny / Wilken, Patrick / Herold, Christian (2020): Flexible Customization of a Single Neural Machine Translation System with Multi-dimensional Metadata Inputs. In: Proceedings of the 14th Conference of the Association for Machine Translation in the Americas. S. 204–216. https://aclanthology.org/2020.amta-user.10/ [28.09.2025].

[2] Dougal, Duane K. / Lonsdale, Deryle (2020): Improving NMT Quality Using Terminology Injection. In: Proceedings of the Twelfth Language Resources and Evaluation Conference. S. 4820–4827. https://aclanthology.org/2020.lrec-1.593/ [27.09.2025].

[3] Schneeberger, Jessica (2024): Gendervarianten in der maschinellen Übersetzung. In: edition. Fachzeitschrift für Terminologie. Ausgabe 2024(1). S. 23–30.

[4] Šorak, Vanessa (im Druck): Neural Machine Translation in Pandemic-Related Health Communication. A comprehensive analysis of risks and risk mitigation strategies using WHO's COVID-19 communication as an example. Veröffentlichung 2026.

[5] Hendy, Amr / Abdelrehim, Mohamed / Sharaf, Amr / Raunak, Vikas / Gabr, Mohamed / Matsushita, Hitokazu / Kim, Young Jin / Afify, Mohamed / Awadalla, Hany (2023): How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation. arXiv. doi:10.48550/arXiv.2302.09210.

[6] Jiang, Zhaokun / Lv, Qianxi / Zhang, Ziyin / Lei, Lei (2024): Convergences and Divergences between Automatic Assessment and Human Evaluation: Insights from Comparing ChatGPT-Generated Translation and Neural Machine Translation. arXiv. doi:10.48550/arXiv.2401.05176.

[7] World Health Organization: Support for rehabilitation: self-management after COVID-19-related illness, second edition, Abschnitt „Managing problems with your voice", S. 16. https://iris.who.int/items/8304692f-dc66-4505-baa1-e28bfd8f44f0 [28.09.2025].

[8] World Health Organization: The prevention and elimination of disrespect and abuse during facility-based childbirth: WHO statement, Abschnitt „Background“, S. 1f. https://iris.who.int/handle/10665/134588 [28.09.2025].
[9] He, Sui (2024): Prompting ChatGPT for Translation: A Comparative Analysis of Translation Brief and Persona Prompts. In: Proceedings of the 25th annual conference of the European association for machine translation. S. 316–326. doi:10.48550/arxiv.2403.00127.
[10] Šorak, Vanessa (2025): Terminologieerzwingung mit generativen LLMs. Datensatz. Verfügbar über Zenodo. doi:10.5281/zenodo.17273631.

Vanessa Šorak ist wissenschaftliche Mitarbeiterin am Institut für Übersetzen und Dolmetschen der Universität Heidelberg. Ihr Forschungsschwerpunkt liegt auf Künstlicher Intelligenz im Kontext maschineller Übersetzung.

Kontaktadresse
vanessa.sorak@iued.uni-heidelberg.de
www.iued.uni-heidelberg.de

Haben professionelle Übersetzungswörterbücher noch eine Existenzberechtigung?

Maurice Mayer-Dewor und Giovanni Rovere

This article examines whether using professional translation dictionaries can still be relevant in modern translation workflows. By comparing machine translations of Italian verbs with dictionary entries of a valency-dictionary of Italian verbs, the Wörterbuch der italienischen Verben, it assesses the limitations of current AI tools in achieving semantically and pragmatically adequate translations.

Keywords: translation dictionaries, machine translation, artificial intelligence, valency theory, verbs

In der *ZEIT* vom 5. September 2024 findet sich folgende Passage: „Auf den kargen Fluren der Brüsseler Übersetzungsabteilung herrscht derzeit eine heitere Untergangsstimmung. Heiter, weil sie [die Übersetzer] wissen, dass ihnen als Beamten keine Entlassung droht. Untergang, weil die KI anfängt, mühelos die Arbeit zu erledigen, für die sie lange studiert haben und einen harten Zulassungstest absolvieren mussten." (S. 12)

Der auf KI-Sprachtechnologien beruhende maschinelle Übersetzer, der von einem menschlichen post-editiert wird, soll also allmählich durch einen maschinellen Übersetzer ersetzt werden, der schließlich autonom arbeiten wird. Diskussionsgegenstand ist vielmals nicht mehr die Frage, ob die Entwicklung diesen Punkt erreichen wird, sondern wann und in welchem Umfang. Dieser Prozess zeichnet sich selbstredend auch in vielen anderen Berufen wie Finanzanalyst, Rechtsassistent oder Grafikdesigner ab, vgl. hierzu z. B. International Monetary Fund, Gen-AI: Artificial Intelligence and the Future of Work [1].

Wir wollen uns hier jedoch nicht mit den Zukunftsperspektiven des Arbeitsmarktes für Übersetzer und Terminologen befassen, sondern zu ergründen versuchen, ob ein professionelles Übersetzungswörterbuch heute noch zu den gängigen Werkzeugen von Übersetzern und Terminologen zählen kann und wenn ja, in welchem Maße. In unserem Fall geht es um das „Wörterbuch der italienischen Verben" (WIV), das in Buchform (Herausgeber P. Blumenthal und G. Rovere) 1998 von Klett in Stuttgart publiziert wurde und seit 2016 unter Mitwirkung von M. Mayer-Dewor in einer elektronischen und zudem aktualisierten sowie stark erweiterten Version innerhalb der Wörterbuchplattform UniLex (Acolada, Nürnberg) – zuletzt 2025 in der 6. Auflage (Herausgeber P. Blumenthal, G. Rovere, M. Mayer-Dewor) – erscheint [2,3]. Ein Zusammenhang zwischen unserer Fragestellung und den Entwicklungen auf dem Arbeitsmarkt besteht indessen insofern, als automatisch erstellte Übersetzungen, die von professionellen Übersetzern als fehlerfrei, d. h., gegenüber dem Ausgangstext als semantisch und pragmatisch adäquat übersetzt bewertet werden, offenkundig nicht den Einsatz weiterer Arbeitstools erfordern. Wir wollen im Folgenden punktuell überprüfen, ob bei Übersetzungen – insbesondere fachsprachlicher Verben – vom Italienischen ins Deutsche im Fall einer MÜ- bzw. KI-basierten Übersetzung dies beim heutigen Entwicklungsstand bereits so sei. Als Vertreter der MÜ-Systeme haben wir DeepL aufgrund seiner Spezifizität als KI-Übersetzungstool und seiner Benutzerfreundlichkeit sowie darüber hinaus die generischen und mächtigeren großen Sprachmodelle ChatGPT, Gemini AI, Perplexity und Claude AI ausgewählt.

Unser Vorgehen besteht zunächst darin, bei einigen Belegen zu im WIV bearbeiteten Verben, deren maschinelle Übersetzungen mit den jeweiligen Vorschlägen zur Übersetzung des Verbs im WIV zu vergleichen [4]. Dabei gehen wir von folgenden Hypothesen aus: Wir erwarten, dass erstens MÜ und KI unzutreffende Übersetzungen liefern, wenn die Wortverbindungen, in denen das Verb auftritt, in den zugrunde liegenden Datensätzen mutmaßlich selten oder gar nicht vorkommen. Unter diesen Bedingungen dürften die ermittelten Muster häufig zu übergeneralisierten Lösungen führen. Diese sind leicht als falsch zu erkennen, da sie zumeist in auffälliger Weise sinnentstellend oder inkohärent sind, vgl. z. B. die Übersetzung 1a:

1. Le telecamere […]. Infallibili nel rilevare gli ingressi abusivi nelle zone a traffico limitato. Utili per punire i furbi. Ma spesso ottusamente incapaci nel distinguerli da chi, per esempio, si affretta a **disimpegnare** (così recita il Codice della strada) un incrocio con semaforo. (Il Sole 24 ORE)

a. Die Kameras […]. Unfehlbar bei der Erkennung von unbefugtem Eindringen in Sperrzonen. Nützlich bei der Bestrafung der Schlauen. Aber oft stumpfsinnig unfähig, sie von denen zu unterscheiden, die z.B. im Eiltempo (so steht es in der Straßenverkehrsordnung) eine Kreuzung mit Ampeln **überqueren**. (DeepL)

b. WIV: **wieder freigeben**

In Verbindung mit *Kreuzung* in Objektposition ist überqueren das häufigste und typischste Verb (vgl. das DWDS-Wortprofil zu *Kreuzung*); *freimachen* und *freigeben*, die hier zutreffenden Lösungen, kommen hingegen erst an 28. und 37. Stelle. In diesem Kontext ergibt jedoch *überqueren* keinen Sinn, da die Vorstellung, das positive Verhalten bestünde in einer raschen Überquerung der Ampelkreuzung, impliziert, dass das gegensätzliche Verhalten der Schlaumeier („i furbi“) notwendigerweise darin bestehen würde, die Kreuzung nicht rasch zu überqueren – ein Unterschied, der von einer Kamera zweifelsfrei erfasst würde. Der Verweis auf die Diktion der Straßenverkehrsordnung bezieht sich nicht auf „si affretta“, sondern auf „disimpegnare“, das im Vergleich zum gleichbedeutenden nichtmarkierten *sgombrare* als formal einzustufen ist. Aus diesem Grund ist im Deutschen das formale Übersetzungsäquivalent *freigeben* gegenüber dem gleichbedeutenden, aber unmarkierten *freimachen* vorzuziehen. In Klammern ist anzumerken, dass auch die Übersetzung von *ingressi abusivi nelle zone a traffico limitato* mit „unbefugtes Eindringen in Sperrzonen“ statt „unbefugtes Einfahren in verkehrsberuhigte Bereiche“ das Verständnis der Übersetzung zusätzlich erschwert.

Unzutreffende Übersetzungen dürften zweitens auftreten, wenn das Verb – zumindest aus der Perspektive der Zielsprache – im ausgangssprachlichen Text Mehrdeutigkeiten oder Unschärfen aufweist. Darauf gründende Fehldeutungen sind in der Regel eher unauffällig, vgl. hierzu den folgenden Abschnitt.

Im Gegensatz zum Vorgehen von MÜ und KI gründet der im WIV gewählte Beschreibungsansatz auf dem Prinzip, dass das Verb die syntaktischen und semantischen Beziehungen zwischen seinen Begleitern bestimmt und dadurch die Bedeutungsanalyse des Satzes steuert. In fachsprachlichen Kontexten, in denen das Verb eine generische Eigenbedeutung hat und die Substantive wichtigste Informationsträger sind, werden die Beziehungen zwischen den informationstragenden Satzkomponenten durch das Verb primär auf syntaktischer Ebene bestimmt. Zwar gelten Substantive als die typischen Träger fachlicher Definitionen und werden demzufolge bevorzugt oder ausschließlich in Fachwörterbüchern und Terminologiesammlungen aufgenommen, in fachsprachlichen Texten lassen sich jedoch häufig Verben finden, die eine definierbare oder im Text definierte fachliche Bedeutung haben und somit auch auf semantischer Ebene einen relevanten Stellenwert in der Strukturierung des Satzes einnehmen können. Ähnlich wie bei der Valenz in der Chemie, bei der es, grob gesprochen, um die Fähigkeit eines Atoms geht, mit einer bestimmten Anzahl anderer Atome eine Verbindung einzugehen, sind im WIV die Satzbaupläne eines Verbs entsprechend der Antwort auf folgende Frage dargestellt: Welche Verbpartner treten in allen Verwendungen eines Satzbauplans auf, sind also obligatorisch, und welche sind zwar fakultativ, kommen aber so häufig und in einer für das Verb typischen Weise vor, dass sie ebenfalls zum entsprechenden Satzbauplan gehören.

Während bei MÜ und KI aus riesigen Datenmengen Muster von Wortverbindungen auf der Grundlage ihrer Wahrscheinlichkeitsverteilung ermittelt werden und in die Übersetzungen eingehen, sucht der Benutzer des WIV in den aufgeführten Satzbauplänen des zu übersetzenden Verbs denjenigen aus, der semantisch und pragmatisch der Verwendung des Verbs im Ausgangstext entspricht. Hierbei orientiert er sich an den jeweiligen lexikographischen Angaben, vgl. z. B. zu *disimpegnare*:

***disimpegnare*, Beispiel-Satzbauplan #1**
Fachbereich: wirts., finanz.;
Subjekt: Institution;
direktes Objekt: Finanzmittel

Belegbeispiel: La Regione punta a **disimpegnare** parte di queste risorse per coprire i debiti pregressi maturati sulla sanità e sul trasporto pubblico locale. (Corriere della Sera) [5]
dt. ***freigeben***

***disimpegnare*, Beispiel-Satzbauplan #2**
Fachbereich: wirts., finanz.;
Subjekt: Unternehmen, Anleger, Sparer;
direktes Objekt: Finanzmittel

Belegbeispiel: Il tetto massimo finanziabile è di 200mila euro ma in certi casi si può arrivare a 300mila euro (ad esempio se il contraente è un libero professionista che per la compravendita non voglia **disimpegnare** denaro investito in precedenza). (Il Sole 24 ORE)
dt. ***freisetzen***

Innerhalb der Satzbaupläne werden dem Benutzer die zu allen aufgeführten Belegen passenden Übersetzungsäquivalente angeboten. *Disimpegnare* weist beispielsweise fünfzehn syntaktisch, semantisch oder pragmatisch unterschiedene Satzbaupläne auf. Sie enthalten insgesamt 113 Belege und Beispiele. Die Belege sind gezielt nach Valenzkriterien aus den insgesamt knapp über 1000, aus den Korpora des WIV automatisch extrahierten ausgewählt. Die Beispiele hingegen sind aus allgemeinen einsprachigen Wörterbüchern entnommen und decken die in der italienischen Lexikographie als typisch eingeschätzten Gebrauchsweisen des Verbs ab. Den Belegen und Beispielen sind im Falle von *disimpegnare* 59 unterschiedliche Übersetzungsäquivalente beigefügt, gegebenenfalls im Rahmen von kurzen

oder ausführlichen kontextuellen Übersetzungen. Der hohe Differenzierungsgrad in der Darstellung des ausgangssprachlichen Verbs ist die Voraussetzung für ein möglichst passgenaues Übersetzungsangebot.

Im Gegensatz zu MÜ und KI, bei denen auch sehr umfangreiche Übersetzungen annähernd zeitgleich mit der Eingabe des entsprechenden Ausgangstextes auf dem Bildschirm erscheinen, ist die Suche im WIV naturgemäß zeitaufwändiger, in Abhängigkeit vor allem vom Komplexitätsgrad der zu übersetzenden Textstelle. Dem quantitativen Unterschied zwischen den den großen Sprachmodellen zugrunde liegenden Datenmengen und den im Vergleich dazu sehr bescheidenen ausgangssprachlichen Korpora des WIV im Umfang von mehr als 10 GByte an reinen Textdaten mit insgesamt rund 2 Mrd. Wörtern steht ein qualitativer gegenüber. Die großen Sprachmodelle leiten aus riesigen Datensätzen, bestehend aus Texten aller Art, die Übersetzung ab, die auf der Grundlage der Mustererkennung als die wahrscheinlichste gilt, das WIV arbeitet hingegen mit Korpora, die gezielt den Bedürfnissen professioneller Übersetzer entsprechend ausgesucht und größtenteils nicht frei im Netz verfügbar sind. Die Auswahl orientiert sich an den hauptsächlichen Fachgebieten des professionellen Übersetzens: Wirtschaft, Medizin, Recht, Technik sowie nachrangig Politik, Religion, Architektur, Freizeit, Sport, Kulinarik. Die Übersetzungen der Belege werden durchweg manuell erstellt.

MÜ und KI

„Verschaffen Sie Ihrem Unternehmen einen Wettbewerbsvorteil mit KI-gestützten Übersetzungen, die selbst die Qualität von Tech-Giganten übertreffen und einen messbaren ROI erzielen.“ (Werbung von DeepL, 2025) Sicher ist, dass DeepL im Vergleich zu den großen Sprachmodellen den praktischen Vorteil einer einfachen Suchmaske bietet, die ein automatisches Abrufen auch umfangreicher Übersetzungen erlaubt. Es ist allerdings zu klären, ob tatsächlich qualitative Unterschiede zwischen DeepL und den mächtigeren, aber nicht ausschließlich zu Übersetzungszwecken konzipierten Sprachmodellen bestehen. Auch dies soll in den vorgenommenen Vergleichstests punktuell überprüft werden.

2. Una tale operazione potrebbe sì permettere alle banche di **disimpegnarsi**, ma non risolverebbe il problema principale. (Il Sole 24 ORE)

a) Eine solche Maßnahme könnte es den Banken ermöglichen, **sich zu entlasten**, würde aber das Hauptproblem nicht lösen. (DeepL)

b) Eine solche Maßnahme könnte den Banken zwar ermöglichen, **sich zu entlasten**, aber sie würde das Hauptproblem nicht lösen. (ChatGPT)

c) Eine solche Operation könnte es den Banken zwar ermöglichen, **sich zu entlasten**, würde aber das Hauptproblem nicht lösen. (Gemini AI)

d) Eine solche Operation könnte es den Banken zwar ermöglichen, **sich zurückzuziehen**, würde aber das Hauptproblem nicht lösen. (Perplexity)

e) Eine solche Operation könnte es den Banken zwar ermöglichen, **sich zurückzuziehen** (sich aus der Verantwortung zu ziehen), würde aber das Hauptproblem nicht lösen. (Claude AI)

f) WIV: ...**ihr Engagement zu beenden**...

Die aktive Konstruktion *die Bank entlasten* ist gut belegt, ebenso *ein Konto entlasten* in der Bedeutung ‚auf einem Konto einen Schuldbetrag tilgen‘. Die seltene reflexive Konstruktion *sich entlasten* ist in psychologischen Kontexten (*sich von innerem Druck entlasten*) und juristischen (*sich von einer Schuld entlasten*) zu finden, nicht jedoch im Bereich des Bankwesens. *Sich zurückziehen* in den identischen Übersetzungen 2d und 2e ist korrekt, allerdings ist der Vorschlag des WIV aus pragmatischen Gründen vorzuziehen, da in 2 *disimpegnarsi* gegenüber häufigerem und gleichbedeutendem *ritirarsi* als gehoben markiert ist.

3. Il tetto massimo finanziabile è di 200mila euro ma in certi casi si può arrivare a 300mila euro (ad esempio se il contraente è un libero professionista che per la compravendita non voglia **disimpegnare** denaro investito in precedenza). (Il Sole 24 ORE)

a) Der Höchstbetrag, der finanziert werden kann, beträgt 200.000 EUR, kann aber in bestimmten Fällen bis zu 300.000 EUR betragen (z. B. wenn der Versicherungsnehmer ein Freiberufler ist, der **sich** nicht von bereits investiertem Geld **trennen** möchte). (DeepL)

b) Der maximale finanzierbare Betrag liegt bei 200.000 Euro, kann in bestimmten Fällen jedoch auf 300.000 Euro steigen (zum Beispiel, wenn der Vertragspartner ein Freiberufler ist, der für den Kauf und Verkauf kein zuvor investiertes Geld **freigeben** möchte). (ChatGPT)

c) Die maximal finanzierbare Obergrenze beträgt 200.000 Euro, kann aber in bestimmten Fällen auf 300.000 Euro steigen (z. B. wenn der Vertragspartner ein Freiberufler ist, der für den Kauf kein zuvor investiertes Geld **freisetzen** möchte). (Gemini AI)

d) Die maximale finanzierbare Obergrenze beträgt 200.000 Euro, aber in bestimmten Fällen kann sie bis zu 300.000 Euro erreichen (zum Beispiel, wenn der Vertragspartner ein Freiberufler ist, der für den Kauf kein zuvor investiertes Geld **freisetzen** möchte). (Perplexity)

e) Die maximale finanzierbare Summe beträgt 200.000 Euro, aber in bestimmten Fällen kann man bis zu 300.000 Euro erreichen (zum Beispiel wenn der Vertragspartner ein Freiberufler ist, der für den Kauf kein zuvor investiertes Geld **freisetzen/freimachen** möchte). (Claude AI)

f) WIV: **freisetzen, freimachen**

Die in Claude AI in Klammern angeführte Übersetzung könnte eine kontextuelle Paraphrase oder ein Alternativvorschlag darstellen oder als angemessenere Lösung zu verstehen sein, falls der erweiterte Kontext diese nahelegt.

Sinnentstellend sind Übersetzung 3a, da es hier nicht um einen Verzicht auf das investierte Geld geht, und 3b, weil in 3 das investierte Geld nicht für andere verfügbar sein soll, sondern für den Betreffenden selbst. *Freigeben* tritt in wirtschafts- und finanzsprachlichen Texten auf, in denen das handelnde Subjekt eine Institution ist. Wie weiter oben dargelegt, sind in den entsprechenden deutschen Texten *freisetzen, freimachen* üblich.

Versicherungsnehmer (statt *Vertragspartner*) als Übersetzung von *contraente* in DeepL geht möglicherweise auf das automatische und systematische Sammeln von Daten aus Wörterbüchern zurück, vgl. den Eintrag zu *contraente* im „Großwörterbuch Italienisch PONS-Zanichelli“: „Vertragspartner(in) m(f), Vertragsschließende m(f) [...], Kontrahent(in) m(f) […]; (*nelle assicurazioni*) Versicherungsnehmer(in) m(f)“.

Compravendita entspricht dem Deutschen *An- und Verkauf*. In 3 bedingt *disimpegnare denaro investito* jedoch, dass die Bedeutung auf ‚Kauf‘ eingeschränkt ist. *Kauf und Verkauf* in 3b und *Geld freigeben für den Verkauf* in 3f sind inkohärent.

4. Lancio lungo nella metà campo biancorossa, un difensore **disimpegna di testa vanificando** l'uscita di Benedetti. (Gazzetta dello Sport)

a) Langer Ball in die Hälfte der Rot-Weißen, der **Kopfball** eines Verteidigers **vereitelt** den **Abschluss** von Benedetti. (DeepL)

b) Ein langer Pass in das weiße und rote Spielfeld, ein Verteidiger **klärt mit dem Kopf** und **vereitelt** das Herauslaufen von Benedetti. (ChatGPT)

c) Langer Ball in die weiße Hälfte des Spielfelds, ein Verteidiger **klärt per Kopf** und **vereitelt** so das Herauslaufen von Benedetti. (Gemini AI)

d) Langer Ball in die weiß-rote Spielfeldhälfte, ein Verteidiger **klärt per Kopf** und **macht** damit den Ausflug von Benedetti **zunichte**. (Perplexity)

e) Langer Ball in die rot-weiße Spielfeldhälfte, ein Verteidiger **klärt per Kopfball** und **macht** damit den Ausflug von Benedetti **zunichte**. (Claude AI)

f) WIV: disimpegna - **klären mit**; …vanificando… - **umsonst** […so dass Benedetti umsonst aus dem Tor läuft.]

Sinnentstellend ist 4a, da *uscita* bei Torszenen auf den Torhüter als handelndes Subjekt zu beziehen ist und nicht auf einen Feldspieler.

Ausflug verbindet sich mit *Torhüter* in Situationen, in denen dieser den Strafraum verlässt und sich ins Mittelfeld wagt.

Sinnentstellend sind 4b, 4c und 4d, da *vereiteln* und *zunichtemachen* eine Konfliktsituation zwischen Torhüter und Verteidiger voraussetzen, in der der Verteidiger den Torhüter absichtlich scheitern lässt. Hier gilt hingegen, dass durch den klärenden Kopfball des Verteidigers die gefährliche Situation vor dem Tor bereinigt wird und der Torhüter daher unnötigerweise aus dem Torraum läuft.

Metà campo biancorossa bezeichnet das Spielfeld einer Mannschaft aufgrund ihrer Vereinsfarben.

5. È nocivo per lo sviluppo economico, per esempio, **discriminare** l'imposizione a vantaggio dei rendimenti dei titoli del debito pubblico. (Il Sole 24 ORE)

a) Es ist der wirtschaftlichen Entwicklung abträglich, wenn beispielsweise die Besteuerung zugunsten der Renditen von Staatsanleihen **diskriminiert** wird. (DeepL)

b) Es ist schädlich für die wirtschaftliche Entwicklung, zum Beispiel die Besteuerung zugunsten der Renditen von Staatsanleihen zu **diskriminieren**. (ChatGPT)

c) Es ist schädlich für die wirtschaftliche Entwicklung, beispielsweise die Besteuerung zugunsten der Renditen von Staatsanleihen zu **diskriminieren**. (Gemini AI)

d) Es ist schädlich für die wirtschaftliche Entwicklung, zum Beispiel die Besteuerung zugunsten der Renditen von Staatsanleihen zu **diskriminieren**. (Perplexity)

e) Es ist schädlich für die wirtschaftliche Entwicklung, zum Beispiel die Besteuerung zugunsten der Renditen von Staatsanleihen zu **diskriminieren**. (Claude AI)

f) WIV: **vernachlässigen**

Das mehrfach polyseme Verb *discriminare* entspricht dem deutschen *diskriminieren* in den Lesarten ‚herabsetzen‘ und ‚ausgrenzen‘, wenn es in Verbindung mit Personen und Bevölkerungsgruppen in Objektposition auftritt. In wirtschaftlichen Kontexten bedeutet *discriminare* ‚vernachlässigen‘, eine Bedeutung, die *diskriminieren* nicht hat. Die Übersetzungen 5a–5e sind im vorgegebenen Kontext unüblich.

Bei fehlerhaften Übersetzungen sind Übereinstimmungen zwischen Sprachmodellen Ausdruck der Anwendung gleicher methodologischer Verfahren.

6. Per l'impiego clinico della misurazione delle resistenze è molto importante poter **discriminare** l'eventuale risposta al broncodilatatore. (Rivista di Immunologia e Allergologia Pediatrica)

a) Für die klinische Anwendung der Widerstandsmessung ist es sehr wichtig, die mögliche Reaktion auf den Bronchodilatator **unterscheiden** zu können. (DeepL)

b) Für den klinischen Einsatz der Messung der Widerstände ist es sehr wichtig, die mögliche Reaktion auf das Bronchodilatator zu **unterscheiden**. (ChatGPT)

c) Für den klinischen Einsatz der Widerstandsmessung ist es sehr wichtig, die mögliche Reaktion auf einen Bronchodilatator **unterscheiden** zu können. (Gemini AI)

d) Für den klinischen Einsatz der Messung des Atemwegswiderstands ist es sehr wichtig, eine mögliche Reaktion auf das Bronchodilatator-Medikament **erkennen** zu können. (Perplexity)

e) Für die klinische Anwendung der Widerstandsmessung ist es sehr wichtig, die mögliche Reaktion auf das Bronchodilatator **unterscheiden/erkennen** zu können. (Claude AI)

f) WIV: **ermitteln, feststellen**

Eine zweite häufig belegte Bedeutung von *discriminare* ist ‚unterscheiden‘. Diese Lesart setzt aus logischen Gründen voraus, dass die Entitäten, die Gegenstand der Unterscheidung sind, eine Mehrzahl bilden (unterschieden wird x von y,

zwischen x und y sowie x im Plural). Im Fall von 6 ist folglich die Übersetzung mit *unterscheiden* syntaktisch fehlerhaft und semantisch inkohärent.

Logische Fehlleistungen dieser Art sind auf die extensionale Begrenztheit des Datenmusters und somit auf die Unfähigkeit zu einer umfassenden Textanalyse zurückzuführen.

In Claude AI werden häufig für die gleiche Textstelle zwei lexikalische Elemente angegeben, die durch Schrägstrich getrennt sind, wobei offenbleibt, ob semantische Unterschiede zwischen ihnen bestehen (und wenn ja welche), ob es sich um Synonymangaben handelt oder ob pragmatische Unterschiede bestehen. Im WIV werden Synonymangaben durch Kommata und semantisch unterschiedliche Lösungen, die eine Polysemie im Ausgangstext anzeigen, durch Semikolon getrennt.

7. Con un linguaggio di sobria ed esemplare chiarezza unito ad un'inesauribile curiosità, questo grande storico dell'architettura **dirama** la sua analisi lungo molteplici itinerari. (L'Indice)

a) Mit einer Sprache von nüchterner und vorbildlicher Klarheit, gepaart mit unerschöpflicher Neugier, **verzweigt** dieser große Architekturhistoriker seine Analyse entlang mehrerer Routen. (DeepL)

b) Mit einer Sprache von nüchterner und beispielhafter Klarheit, verbunden mit einer unerschöpflichen Neugier, **entfaltet** dieser große Architekturhistoriker seine Analyse entlang vielerlei Wege. (ChatGPT)

c) Mit einer Sprache von nüchterner und beispielhafter Klarheit, gepaart mit unerschöpflicher Neugier, **verzweigt** dieser große Architekturhistoriker seine Analyse entlang vielfältiger Routen. (Gemini AI)

d) Mit einer Sprache von nüchterner und beispielhafter Klarheit, gepaart mit einer unerschöpflichen Neugier, **verzweigt** dieser große Architekturhistoriker seine Analyse entlang vielfältiger Wege. (Perplexity)

e) Mit einer Sprache von nüchterner und beispielhafter Klarheit, gepaart mit einer unerschöpflichen Neugier, **verzweigt** dieser große Architekturhistoriker seine Analyse entlang vielfältiger Wege. (Claude AI)

f) WIV: ...**entfaltet** seine Analyse auf vielfältigen Pfaden., ...**entfaltet** seine Analyse entlang vielfältiger Pfade.

Die Übersetzung von *diramare* mit *verzweigen* in 7a, 7c, 7d und 7e scheint durch die wesentlich häufigere Lesart von reflexivem *diramarsi* als Zustandsverb mit *Weg, Fluss, Ast* u. Ä. in Subjektposition und der Bedeutung ‚sich verzweigen' bestimmt zu sein, vgl. den Kommentar in Perplexity: „*dirama la sua analisi* wird als *verzweigt seine Analyse* übersetzt. Das Verb *verzweigen* gibt das italienische *diramare* gut wieder." Die Anmerkung dürfte eine Bestätigung der vielfach zu beobachtenden Tendenz zur Generalisierung der am häufigsten belegten Bedeutung eines Verbs darstellen. In 7 wird das Verb hingegen offenkundig in übertragener Bedeutung verwendet. Im Deutschen verbindet sich *verzweigen* in übertragener Verwendung nur mit Wörtern wie *Analyse* u. Ä. in Subjektposition (*die Analyse verzweigt sich*).

Das WIV wählt *Pfad* und nicht *Weg*, da *Pfad* besser die übertragene Bedeutung von *itinerario* ‚Reihe von Schritten, die unternommen werden müssen, um ein Ziel zu erreichen' wiedergibt.

8. Nel postoperatorio, i pazienti **sono stati allettati** per circa 48 ore in base al livello emoglobinico ed allo stato generale. (Giornale Italiano di Ortopedia e Traumatologia)

a) In der postoperativen Phase **waren** die Patienten je nach Hämoglobinwert und Allgemeinzustand für etwa 48 Stunden **bettlägerig**. (DeepL)

b) Im postoperative (sic) Zeitraum wurden die Patienten je nach Hämoglobinwert und allgemeinem Zustand für etwa 48 Stunden **bettlägerig gehalten**. (ChatGPT)

c) Nach der Operation wurden die Patienten je nach Hämoglobinwert und Allgemeinzustand etwa 48 Stunden lang **bettlägerig gehalten**. (Gemini AI)

d) In der postoperativen Phase wurden die Patienten je nach Hämoglobinspiegel und Allgemeinzustand für etwa 48 Stunden **bettlägerig gehalten**. (Perplexity)

e) Im postoperativen Zeitraum wurden die Patienten für etwa 48 Stunden **bettlägerig gehalten**, abhängig vom Hämoglobinspiegel und dem Allgemeinzustand. (Claude AI)

f) WIV: ...wurde den Patienten **Bettruhe** für 48 Stunden...**verordnet**.

9. Non si corre così il rischio di **allettare** all'impegno politico soltanto chi non abbia migliori alternative. (Repubblica)

a) Dies birgt nicht die Gefahr, nur diejenigen zum politischen Engagement zu **verleiten**, die keine besseren Alternativen haben. (DeepL)

b) So besteht nicht die Gefahr, dass nur diejenigen für das politische Engagement **gewonnen** werden, die keine besseren Alternativen haben. (ChatGtp)

c) So besteht nicht die Gefahr, dass nur diejenigen zum politischen Engagement **verleitet** werden, die keine besseren Alternativen haben. (Gemini AI)

d) Man läuft so nicht Gefahr, nur diejenigen zum politischen Engagement zu **verlocken**, die keine besseren Alternativen haben. (Perplexity)

e) Auf diese Weise läuft man nicht Gefahr, nur diejenigen zur politischen Betätigung zu **verlocken**, die keine besseren Alternativen haben. (Claude AI)

f) WIV: **bewegen zu**

Dass es sich hier bei *allettare* in 8 und 9 um zwei gleichlautende Wörter mit ganz unterschiedlichen Bedeutungen handelt, also ein Fall von Homonymie vorliegt, ist in allen Übersetzungen erkennbar. Die Passivkonstruktion in 8 stellt eine ärztliche Maßnahme in den Mittelpunkt, wobei die Person, die die Maßnahme verordnet, der Arzt, im Rahmen eines Krankenhauses als handelnde Person vorgegeben ist und somit in der Satzkonstruktion prinzipiell ausgelassen wird. Die Übersetzung 8a hebt hingegen den Zustand der Patienten hervor, die erkrankt zu Bett liegen. Die Wendung *bettlägerig halten*, die in 8b, 8c und 8e gewählt wird, verweist in der Regel auf die meist negativ dargestellte Situation von Heimbewohnern, die vom Pflegepersonal aus

vornehmlich nichtmedizinischen Gründen wie Überlastung u. Ä. im Bett gehalten werden.

Politisches Engagement (9) ist ein Verhalten, das positiv bewertet wird. Dem entsprechen *gewinnen für* in 9b und *bewegen zu* (WIV). *Verleiten zu* (‚jmdn. dazu bringen, etwas Falsches zu tun') in 9a und *verlocken* (‚jmdn. zu etw. verführen') in 9d und 9e sind folglich sinnentstellend.

10. Dicono che quando si pesca ci si **astrae** totalmente. (Il Sole 24 ORE)

a) Man sagt, dass man sich beim Angeln völlig **abstrahiert**. (DeepL)

b) Man sagt, dass man beim Angeln völlig **abschaltet**. (ChatGPT)

c) Man sagt, beim Angeln **schaltet** man komplett **ab**. (Gemini AI)

d) Man sagt, dass man beim Angeln völlig **abschaltet**. (Perplexity)

e) Sie sagen, dass man beim Angeln völlig in eine andere Welt **eintaucht**. (Claude AI)

f) WIV: ...beim Angeln völlig **abschaltet**.; ...sich beim Angeln völlig **entspannt**.

Astrarre hat in den meisten Konstruktionsweisen *abstrahieren* als Übersetzungsäquivalent. Dies mag erklären, weswegen es auch in 10a eingesetzt wird. Laut GWDS (Duden – Das große Wörterbuch der deutschen Sprache) bedeutet *sich abstrahieren* jedoch ‚von sich absehen' und ‚auf etwas verzichten' und ist als bildungssprachlich markiert. *In eine andere Welt eintauchen* in 10e ist sinnenstellend. Die sentenzartige Formulierung in 10 legt nahe, dass *dicono* unpersönlich ist, vgl. hingegen 10e. Das WIV gibt zwei Übersetzungsvorschläge, die sich pragmatisch unterscheiden: *abschalten* ist im Gegensatz zu *sich entspannen* umgangssprachlich.

Eine Zwischenbilanz

Die im Folgenden gezogene Bilanz ist aus zwei Gründen als vorläufig zu betrachten. Zum einen steht zu erwarten, dass die Sprachmodelle, da sie von ihrer Konzeption her einem kontinuierlichen Erweiterungsprozess unterzogen sind, ihre Leistungsfähigkeit ebenfalls kontinuierlich steigern. Zum andern ist unsere Untersuchung, die auf die Verwendung ausgewählter Verben in ausgewählten Textstellen begrenzt ist, durch einen systematischen Ansatz zu erweitern, in dem die Übersetzungen von mindestens einem Verb in allen seinen Konstruktionen erfasst werden.

DeepL hat seine Vorzüge in der bekanntlich unkomplizierten Benutzbarkeit. Der Vergleich zwischen den ausgewählten maschinellen Übersetzern zeigt, dass die Qualität der Übersetzungen jedoch, insgesamt betrachtet, leicht geringer zu sein scheint als die der berücksichtigten großen Sprachmodelle. Unter diesen sticht – anders als möglicherweise zu erwarten war – keines als das eindeutig beste hervor.

Gegenüber DeepL haben die großen Sprachmodelle den Vorzug, dass sie häufig Erläuterungen zum Ausgangstext und zur Übersetzung der einzelnen Textkomponenten geben, automatisch oder als Suchergebnis. Zudem lassen sich die Ergebnisse durch Prompts beeinflussen. Nachteilig ist das Angebot unterschiedlicher Übersetzungsangaben, wenn deren semantischer Status nicht kommentiert wird.

Im Vergleich zu einem Wörterbuch haben alle maschinellen Übersetzer den offenkundigen Vorteil, dass sie auch längere Textpassagen mit hoher Geschwindigkeit übersetzen. Was die Qualität der Übersetzungen angeht, fällt die Bilanz ernüchternd aus: Für Übersetzungen, die sich nicht auf zumindest teilstandardisierte Texte beschränken, sondern hohen Ansprüchen an semantischer und pragmatischer Adäquatheit gerecht werden müssen, erweisen sich die berücksichtigten Tools als unzureichend. Unter diesen Bedingungen stellt sich die verbreitete Vorstellung, der Übersetzer brauche, falls überhaupt, nur noch leicht korrigierend einzugreifen, allgemein betrachtet – zum aktuellen Zeitpunkt – als unrealistisch heraus.

Quellen und Endnoten

[1] Cazzaniga, Mauro et al. (2024): "Gen-AI: Artificial Intelligence and the Future of Work". In: Staff Discussion Notes, Volume 2024: Issue 001. International Monetary Fund. https://doi.org/10.5089/9798400262548.006

[2] Projekt-Website des WIV (Wörterbuch der italienischen Verben): https://www.uni-heidelberg.de/wiv

[3] Produkt-Website des WIV: https://acolada.de/woerterbuecher/woerterbuch-der-italienischen-verben/

[4] Letzte Konsultation der jeweiligen Quellen: 15.4.2025. Die Abfrage der Übersetzungen in den großen Sprachmodellen erfolgte anhand von einfachen Abfragen (One-Shot-Prompts) über die kostenlos verfügbaren Nutzungsmöglichkeiten in der jeweils aktuellen Version.

[5] Da in diesem Beitrag die Verbvalenz im Fokus steht, wird auf eine vollständige Übersetzung der Belege verzichtet.

Maurice Mayer-Dewor ist Diplom-Übersetzer für die Sprachen Italienisch und Französisch und arbeitet als wissenschaftlicher Dozent am Institut für Übersetzen und Dolmetschen (IÜD) der Universität Heidelberg, wo er u. a. im Bereich der Übersetzungstechnologien lehrt.

Kontaktadresse
maurice.mayer@iued.uni-heidelberg.de
www.iued.uni-heidelberg.de

Giovanni Rovere war von 1974 bis 1982 Dozent für italienische Philologie an der Universität Basel, von 1983 bis 2016 Professor für italienische Sprach- und Übersetzungswissenschaft an der Universität Heidelberg. Sein Hauptinteresse gilt der italienischen Rechtssprache und der (Meta-)Lexikographie.

Kontaktadresse
giovanni.rovere@iued.uni-heidelberg.de
www.iued.uni-heidelberg.de

Neue Norm in den Startlöchern

DIN 19460 „Mehrsprachige Terminologiearbeit – Grundsätze und Methoden" erscheint voraussichtlich im Frühjahr 2026

Petra Drewer

Was hat denn Terminologie mit Normung zu tun? Und was wollt ihr da eigentlich normen?
Diese und ähnliche Fragen begegnen mir oft – und nicht selten wird angenommen, dass wir in den DIN-Normen zur Terminologie Fachwörter festlegen und als verbindliche „Sprachregelungen" veröffentlichen.

Die Wirklichkeit ist komplexer – und deutlich spannender. Bevor wir auf die neue DIN 19460 schauen, lohnt sich ein Blick auf die grundlegenden Beziehungen zwischen **Terminologiearbeit** und **Normung**.

Die vielfältigen Verbindungen von Terminologie und Normung

Normung ohne terminologische Klarheit ist kaum denkbar. Egal ob es um die Dichtheit von Schutzanzügen gegen Pflanzenschutzmittel, die Größenberechnung von Damen- und Herrenschuhwerk oder die Sicherheit von Elektrofahrrädern geht: Jede Sachnorm enthält in der Regel ein eigenes Kapitel zur **Begriffsbestimmung** (Kapitel 3). In DIN-Normen trägt dieses Kapitel 3 die Überschrift „Begriffe", in ISO-Normen heißt es „Terms and Definitions". Diese Kapitel sind keine bloßen Formalien. Sie schaffen die Grundlage für das Verständnis und die korrekte Anwendung der Norm – und damit für ihre Wirksamkeit in der Praxis.

Die Ergebnisse solcher Festlegungen finden sich nicht nur in den jeweiligen Normtexten, sondern stehen auch für die Öffentlichkeit zur Verfügung: Das **DIN-TERMinologieportal** (www.din.de/de/service-fuer-anwender/din-term) enthält über eine Million Einträge mit Definitionen aus dem deutschen Normenwerk sowie den entsprechenden europäischen und internationalen Paralleldokumenten und kann online (kostenfrei) eingesehen werden. Normung trägt somit nicht nur zur technischen Standardisierung bei, sondern auch zur Klärung und – wo sinnvoll – Vereinheitlichung der Fachsprache.

Neben den Sachnormen mit Terminologieklärung gibt es terminologische Metanormen, die sich mit den Grundlagen und Methoden der Terminologiearbeit selbst befassen. Zuständig hierfür ist bei DIN der Normenausschuss Terminologie (NAT), der unter anderem Leitlinien zur Erarbeitung und Dokumentation von Begriffen und Benennungen (DIN 2330) oder zu den Arten von Begriffssystemen und zum Umgang damit (DIN 2331) erarbeitet. Diese „**terminologische Grundsatznormung**" schafft die methodischen Grundlagen, auf die sich einerseits andere Normungsprojekte stützen können, andererseits liefert sie ein hilfreiches Fundament für jede Art von Terminologiearbeit, z. B. in Unternehmen und Institutionen, die mit Hilfe von terminologischen Vorgaben eine Corporate Language etablieren möchten.

Dieses komplexe Zusammenspiel zeigt: Terminologiearbeit ist sowohl Voraussetzung für als auch Ergebnis von Normung.

DIN 19460 – ein neues Kapitel der terminologischen Grundsatznormung

Ein aktuelles Beispiel für terminologische Grundsatznormung ist DIN 19460 „Mehrsprachige Terminologiearbeit – Grundsätze und Methoden". Das Dokument wird voraussichtlich 2026 erscheinen. Es „legt Anforderungen für die mehrsprachige Terminologiearbeit fest und gibt Empfehlungen zur Lösung konkreter Probleme, z. B. bei der Verwaltung der mehrsprachigen Terminologie und der Weiterverwendung der terminologischen Daten in anderen Systemen. Es richtet sich an alle, die in Unternehmen und Organisationen an mehrsprachigen Terminologieprozessen beteiligt sind." (E DIN 19460:2025-06, S. 5).

Was ausdrücklich nicht im Vordergrund der Norm steht, sind Übersetzungsprozesse und -theorien, Grundlagen und Methoden der maschinellen Übersetzung oder vollständige Prozesse der Sprachprüfung. Diese Themen würden den Rahmen sprengen und sind bewusst nicht Teil des Anwendungsbereichs der Norm (vgl. E DIN 19460:2025-06, S. 5).

Die Entstehung – Konsens statt Elfenbeinturm

Erarbeitet wurde die neue Norm im Arbeitsausschuss AA 01 „Grundlagen der Terminologiearbeit" des Normenausschusses Terminologie (DIN-NAT). Sie wurde im Juni 2025 als Entwurf veröffentlicht und soll 2026 (nach Einarbeitung aller Kommentare und Einsprüche) als Norm erscheinen, vgl. dazu auch Abbildung 2 „Timeline".

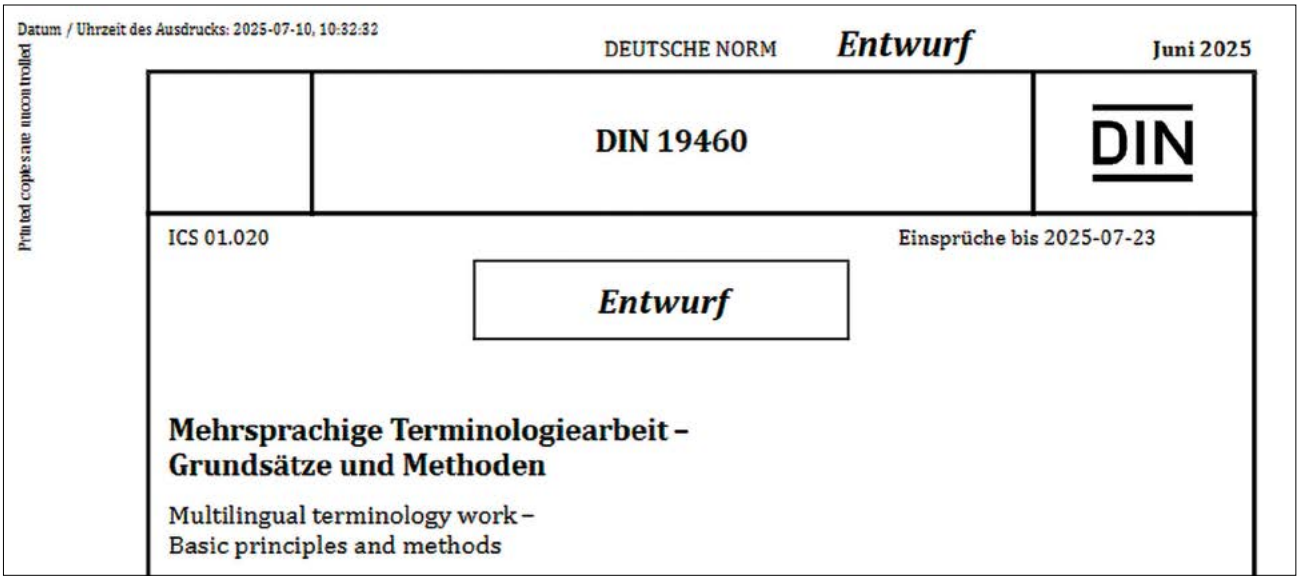


Abb. 1: Deckblatt der E DIN 19460 (Quelle: DIN Media)

Wer noch nie mit Normung zu tun hatte, stellt sich manchmal vor, dass sich eine kleine Gruppe von Expertinnen und Experten bei DIN oder ISO zusammensetzt und nach kurzer Absprache „schnell mal" die wichtigsten Punkte niederschreibt. In Wirklichkeit ist Normungsarbeit ein demokratischer Prozess, geprägt von unterschiedlichen Interessen, lebhaften Diskussionen und dem ständigen Ringen um Konsens.

Auch bei DIN 19460 gab es intensive Abstimmungsprozesse zwischen verschiedenen Personen und Interessengruppen. Konkret waren drei große Gruppen beteiligt, und zwar Vertreterinnen und Vertreter

a) der **Wirtschaft** (u. a. Trumpf Werkzeugmaschinen, Mercedes Benz, Endress+Hauser, ZF Friedrichshafen, Dienstleistungsunternehmen und Softwarehersteller aus den Bereichen Technische Redaktion, Übersetzung, Terminologie)
b) der **öffentlichen Hand** (u. a. Bundessprachenamt, Auswärtiges Amt, Bundeswehr)
c) der **Wissenschaft und Forschung** (u. a. Hochschule Karlsruhe, Goethe-Universität Frankfurt, TH Köln, Universität Mainz)

Gerade in der mehrsprachigen Terminologiearbeit ist eine solide theoretische Basis unverzichtbar. Die Anforderungen sind hoch, einfache Lösungen greifen oft zu kurz. DIN 19460 verbindet daher methodische Strenge mit praxisorientierten Leitlinien – ein Ansatz, der Missverständnisse und Fehlerquellen von Anfang an vermeiden soll. Die Methoden wurden dabei gleichermaßen aus der Fachpraxis wie aus der wissenschaftlichen Reflexion heraus entwickelt.

Normen haben in diesem Zusammenhang eine besondere Stärke – und gleichzeitig eine besondere Schwäche: Sie sind kurz und knapp. Während Fachbücher, wie z. B. Arntz/Picht/Schmitz (2021) oder Drewer/Schmitz (2017), Grundlagen ausführlich darstellen und anhand praxisnaher Beispiele erläutern, oder Tagungsbände, wie z. B. Drewer/Mayer/Pulitano (2023, 2025), aktuelle Entwicklungen aufzeigen und vertiefen, bieten Normen einen kompakten und allgemein akzeptierten Referenzrahmen.

Der Schritt in die Normung eröffnet also eine Zielgruppe, die man sonst vermutlich nicht erreichen würde, wie das folgende Zitat eines Industrievertreters zeigt: *„Für dicke Fachbücher habe ich meist keine Zeit und wissenschaftliche Paper sind oft zu kompliziert, aber die Normen aus meinem Themengebiet kenne ich alle."*

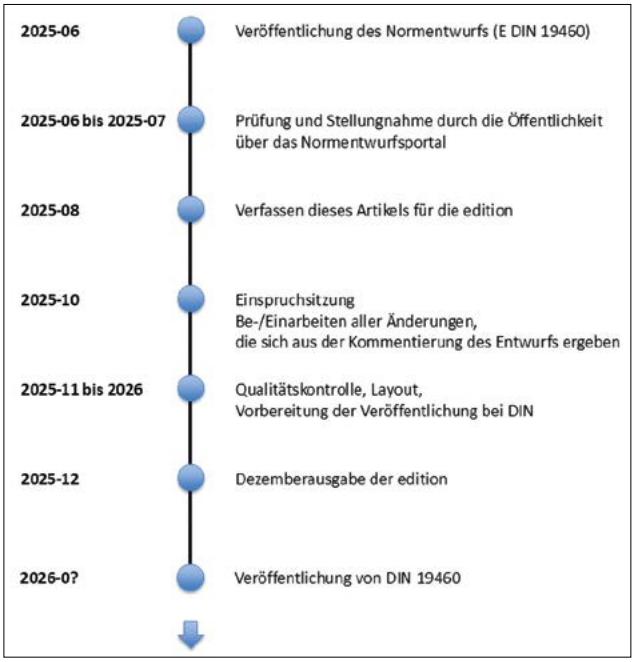


Abb. 2: Timeline zur Veröffentlichung der DIN 19460 (plus Einordnung dieses Artikels) (Quelle: eigene Darstellung)

Was sind die inhaltlichen Schwerpunkte der neuen Norm?

Grundlagen

DIN 19460 beginnt mit einer Einordnung der mehrsprachigen Terminologiearbeit und beschreibt ihre grundsätzliche Methodik. Diese ist besonders anspruchsvoll, da die Begriffe und Begriffssysteme verschiedener Sprach- und Kulturgemeinschaften erarbeitet und verglichen werden müssen. Dazu wird die Methodik der einsprachigen Terminologiearbeit (vgl. DIN 2330), idealerweise unter Einbezug von Muttersprachlern der relevanten Sprachen, auf jede Einzelsprache angewendet:

1. Zunächst wird Sprache A vollständig terminologisch bearbeitet.
2. Anschließend erfolgt die unabhängige Bearbeitung von Sprache B.
3. Erst danach werden A und B miteinander verglichen, um Äquivalenzprobleme zu erkennen und zu lösen.

Durch diese Vorgehensweise kann die typische Überbewertung von Sprache A, d. h. das Unterstellen begrifflicher Identität in allen Sprachräumen, verhindert werden.

Welche anderen methodischen Fehler sind typisch für die mehrsprachige Terminologiearbeit? Oft findet man eine Orientierung an der Benennung (statt am Begriff). Dadurch wird bspw. als Äquivalent für die französische Benennung „installation industrielle“ das deutsche „Industrieinstallation“ angenommen, obwohl es „Industrieanlage“ heißen müsste. Darüber hinaus werden oft Übersetzungen als Recherchegrundlage verwendet oder eine benennungsorientierte Terminologieverwaltung umgesetzt, die sich am Gedanken einer Vokabelliste orientiert, statt den Begriff als Strukturkriterium zu verwenden. Diese und andere Fehler möchte die Norm vermeiden, indem sie die Hintergründe erklärt und Empfehlungen ausspricht.

Äquivalenz

Das Kapitel zur Äquivalenz ist ein Herzstück der Norm. Es umfasst:

- **Allgemeines zu Äquivalenz**
- **Äquivalenzgrade** mit den Unterpunkten
  - Volläquivalenz
    (Eine kleine Anekdote zur Diskussion um den Terminus „Volläquivalenz“ findet sich im Infokasten auf der folgenden Seite.)
  - Teiläquivalenz mit den Unterarten Überschneidung und Inklusion
  - Terminologische Lücke mit den Unterarten Begriffslücke und Benennungslücke
  - Sonderfall „Falsche Freunde“

Zur Verdeutlichung der Äquivalenzgrade hat das Gremium eine Darstellungsform wieder aufgegriffen, die Helmut Felber bereits 1984 genutzt hat, die aber auch heute noch auf eingängige Art und Weise die verschiedenen Äquivalenzgrade visualisiert.

Abbildung 3 in diesem Artikel zeigt die ersten beiden Arten von Äquivalenz in dieser Darstellungsform: Volläquivalenz und Überschneidung (als ein Fall von Teiläquivalenz). Die Kleinbuchstaben repräsentieren jeweils die Begriffsmerkmale, die Großbuchstaben die Begriffsinhalte. Sprache A

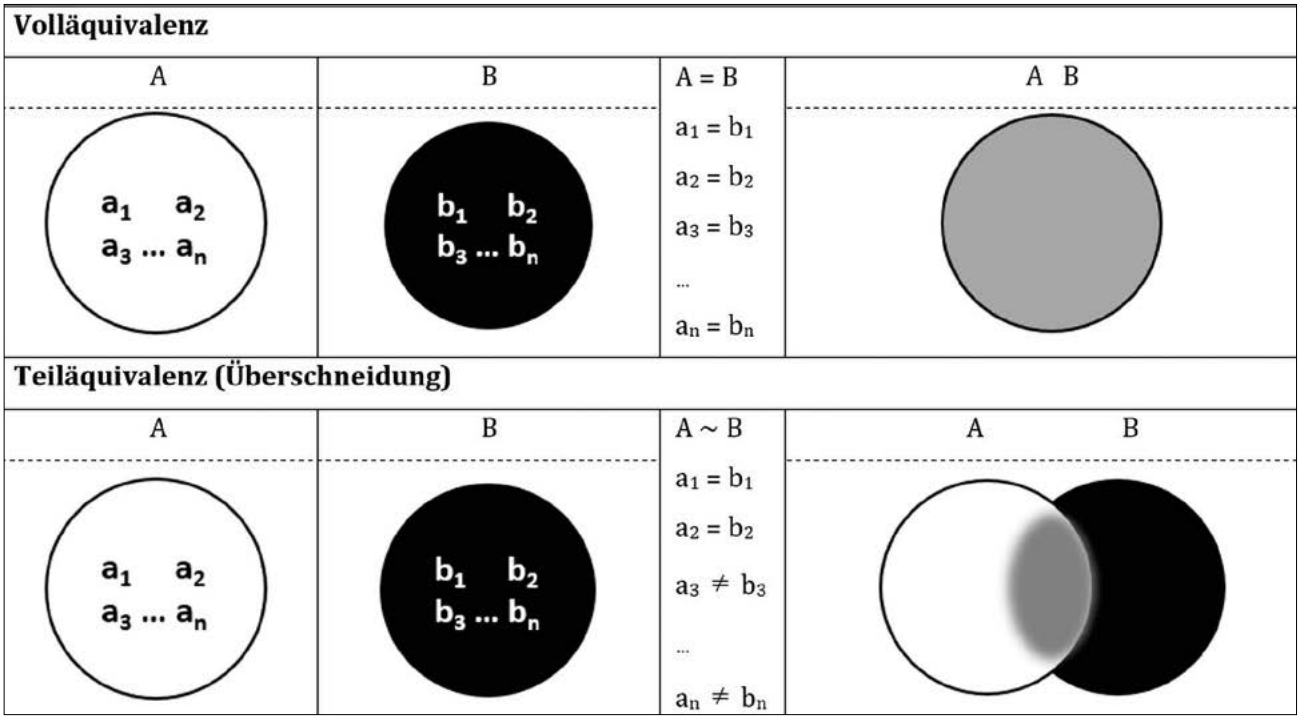


Abb. 3: Bildhafte Darstellung von Äquivalenzgraden in Anlehnung an Felber (1984) (Quelle: eigene Darstellung)

(weiß dargestellt) und Sprache B (schwarz dargestellt) werden zunächst nebeneinander und dann in der rechten Spalte übereinander dargestellt, so dass die Deckungsgleichheit der Begriffe bei der Volläquivalenz ebenso gut zu sehen ist wie die begriffliche Schnittmenge (Überschneidung) bei der Teiläquivalenz. Im ersten Fall sind alle Begriffsmerkmale identisch, im zweiten Fall sind einige identisch, aber jeder Begriff verfügt zusätzlich auch über eigene/spezifische Merkmale.

Definitionen

Definitionen sind das sprachliche Tor zu den Begriffen. Die Norm behandelt daher u. a.:

- die grundsätzliche Bedeutung und besondere Relevanz von Definitionen für die mehrsprachige Terminologiearbeit
- die Auswahl der Sprachen, in denen Definitionen in einer Terminologiedatenbank bereitgestellt werden
- die Position der Definition im Eintragsmodell einer Terminologiedatenbank

Benennungen

Ein umfangreicher Abschnitt widmet sich den Benennungen als sprachlichen Begriffsrepräsentanten, gegliedert in mehrere Unterthemen:

- **Allgemeines:** Regeln zur Bildung und Auswahl von Benennungen sind im Normalfall sprachspezifisch. Sie müssen dokumentiert und allen Beteiligten zugänglich gemacht werden.
- **Sprachen und Varietäten:** Organisationen müssen entscheiden, welche Sprachen und ggf. welche Varietäten (z. B. britisches vs. US-amerikanisches Englisch) in der Terminologiedatenbank verwaltet werden. Die Norm beschreibt verschiedene Möglichkeiten der Verwaltung solcher Varietäten und bewertet sie.
- **Vorzugsbenennungen:** Existieren mehrere Benennungen für denselben Begriff, werden im Rahmen der präskriptiven Terminologiearbeit Vorzugsbenennungen festgelegt. Die Auswahlkriterien können von Sprache zu Sprache variieren und dürfen keinesfalls auf bloßer sprachlicher Ähnlichkeit beruhen.
- **Hoheitsrechte an Begriffen und Benennungen:** Bei Produkt- und Markennamen ist zu prüfen, ob diese in allen Sprachräumen und in anderen Schriftsystemen einsetzbar sind und keine negativen Konnotationen haben.
- **Lösungen für terminologische Lücken:** Die Norm beschreibt Strategien für den Umgang mit Begriffs- und Benennungslücken.

Infokasten: Kleine Anekdote zur Diskussion um den Terminus „Volläquivalenz"

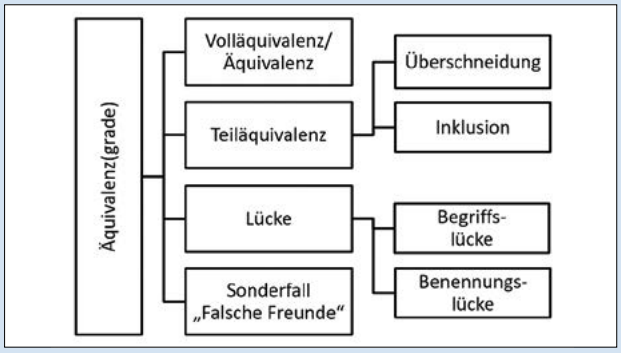


Abb. 4: Äquivalenzgrade nach DIN 19460 (eig. Darstellung)

Wie im Artikel beschrieben, war es unser Ziel, in der Norm vier unterschiedliche Äquivalenzgrade darzustellen (siehe Abb. 4). Im Verlauf der Diskussion stellten wir jedoch fest, dass der Terminus *Äquivalenz* lediglich die vollständige begriffliche Übereinstimmung bezeichnet. Die aktuelle Definition nach DIN 2342:2022-07 [9] (meine Hervorhebung) lautet:

> *Äquivalenz = Beziehung zwischen Bezeichnungen in verschiedenen Sprachräumen* ***für denselben Begriff***
>
> *BEISPIEL*
> *de: serielle Schnittstelle, en: serial interface, fr: interface série*

Diese Verwendung deckt sich mit der in der übersetzungswissenschaftlichen Fachliteratur. Der Begriff *Volläquivalenz* hingegen wird dort nur selten verwendet.

Für unsere Darstellung fehlte uns also ein Oberbegriff, der jede Form von begrifflicher Übereinstimmung umfasst – unabhängig davon, ob sie vollständig oder nur teilweise vorliegt. Mit anderen Worten: Wir brauchten eine klare terminologische Trennung zwischen Äquivalenz im weiteren Sinn und Volläquivalenz als einer spezifischen Ausprägung.

Hätten wir uns strikt an DIN 2342 gehalten (immerhin eine Norm, die auch aus unserem Arbeitsausschuss stammt …), hätte die Einteilung folgendermaßen ausgesehen:

Arten der Äquivalenz

1. Äquivalenz
2. Teiläquivalenz mit den Unterarten Überschneidung und Inklusion
3. Terminologische Lücke mit den Unterarten Begriffslücke und Benennungslücke
4. Sonderfall „Falsche Freunde"

Die Ambiguität von „Äquivalenz" war schwer zu bewältigen, und wir hatten die Befürchtung, dass sie zu vielen Missverständnissen führen würde. Gleichzeitig konnten wir natürlich nicht den etablierten Sprachgebrauch ignorieren.

Nach vielen Diskussionen und Recherchen haben wir uns letztlich dafür entschieden, die Benennung „Volläquivalenz" als Synonym zu „Äquivalenz" aufzunehmen und im Übrigen der in Abbildung 4 dargestellten terminologischen Systematik zu folgen. Die Erweiterung der bestehenden Definition kann man natürlich in Kapitel 3 der neuen Norm sehen, denn wie schon im Artikel gesagt: Kapitel 3 dient der Begriffsklärung …

Datenmodellierung

Dieses Kapitel der Norm beschreibt, wie sich die klassischen Prinzipien der **Begriffsorientierung** und **Benennungsautonomie** in der mehrsprachigen Terminologiearbeit umsetzen lassen. Behandelt werden darüber hinaus insbesondere folgende Aspekte der Datenmodellierung – jeweils illustriert mit verschiedenen Beispielen:

- Auswahl und Anordnung sinnvoller Datenkategorien
- Besonderheiten bei der Verwaltung sprachlicher Varietäten (siehe auch oben im Absatz „Benennungen")
- Erkennen und Verwalten von Ambiguitäten
  Zu diesem Thema liefert der Anhang der Norm ein ausführlich kommentiertes Beispiel, und zwar anhand der ambigen Benennung „Kupplung".
- Modellierung und Verwaltung von Begriffsbeziehungen
- Umgang mit Medien innerhalb der Terminologiedatenbank
- Datenaustausch und Schnittstellen zu anderen Systemen

Anhang mit Praxisbeispielen

Nicht nur ein Anhängsel, sondern ein hilfreicher Bestandteil der Norm ist der Anhang, der mit Hilfe eingängiger zweisprachiger Beispiele typische Äquivalenzprobleme zeigt, ihre Auswirkungen beschreibt und mögliche Lösungen mit Vor- und Nachteilen darstellt.

Fazit

DIN 19460 schließt eine Lücke im Normenwerk: Erstmals gibt es eine spezifische Norm zur mehrsprachigen Terminologiearbeit.

In einer Welt, in der Unternehmen international agieren, Produkte global vermarktet und Fachinformationen maschinell verarbeitet werden, reicht es nicht, Begriffe und Benennungen nur in einer Sprache zu definieren, festzulegen und zu verwalten. Gefragt sind durchdachte, sprachlich abgestimmte und konzeptuell konsistente Terminologien – über Sprachgrenzen hinweg. DIN 19460 will bei der Bewältigung dieser Herausforderungen einen Beitrag leisten.

Literatur

[1] Arntz, Reiner / Picht, Heribert / Schmitz, Klaus-Dirk (2021): Einführung in die Terminologiearbeit. 8. Auflage. Hildesheim: Georg Olms Verlag

[2] DIN 2330 (2022-07): Terminologiearbeit: Grundsätze und Methoden. Berlin: DIN Media

[3] DIN 2331 (2019-12): Begriffssysteme und ihre Darstellung. Berlin: DIN Media

[4] Drewer, Petra / Schmitz, Klaus-Dirk (2017): Terminologiemanagement: Grundlagen – Methoden – Werkzeuge. Berlin/Heidelberg: Springer Vieweg (Kommunikation und Medienmanagement 1)

[5] Drewer, Petra / Mayer, Felix / Pulitano, Donatella (Hrsg.) (2023): Terminologie: Tools und Technologien. Akten des DTT-Symposions 2023. München/Karlsruhe/Bern: Deutscher Terminologie-Tag / SDK

[6] Drewer, Petra / Mayer, Felix / Pulitano, Donatella (Hrsg.) (2025): Terminologie in der KI – KI in der Terminologie. Akten des DTT-Symposions 2025. Karlsruhe/München/Bern: Deutscher Terminologie-Tag / SDK

[7] E DIN 19460 (2025-06): Mehrsprachige Terminologiearbeit – Grundsätze und Methoden. Berlin: DIN Media

[8] Felber, Helmut (1984): Terminology Manual. Paris: Unesco/Infoterm

[9] DIN 2342 (2022-07): Terminologiewissenschaft und Terminologiearbeit – Begriffe. Berlin: DIN Media

Mitarbeit bei DIN

Falls Sie durch diesen Artikel auf den Geschmack gekommen sind: Der Normenausschuss für Terminologie freut sich über Unterstützung in Form von aktiver Mitarbeit.

Infos zum NAT:
https://www.din.de/de/mitwirken/normenausschuesse/nat/ueber-nat

Machen Sie mit im NAT!
https://www.din.de/de/mitwirken/mitarbeit-in-der-normung

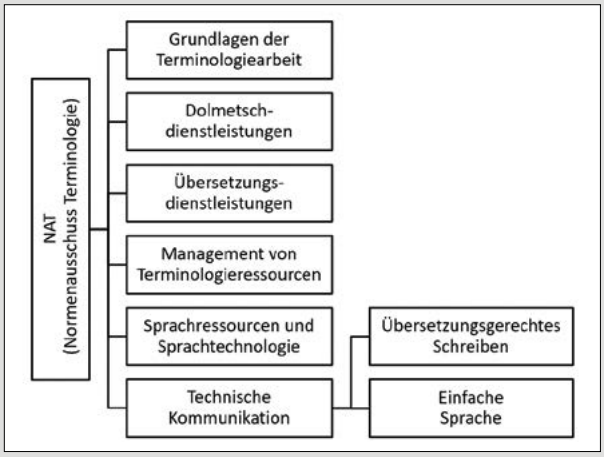
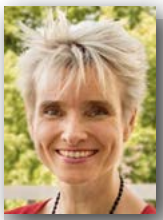


Abb. 5: NAT – Arbeitsausschüsse und Arbeitskreise (Quelle: eigene Darstellung)



Prof. Dr. Petra Drewer ist Studiendekanin im Studiengang „Kommunikation und Medienmanagement" und Prodekanin der Fakultät „Informationsmanagement und Medien" mit den Lehr- und Forschungsschwerpunkten „Sprach- und Terminologiemanagement" der Hochschule Karlsruhe. Sie ist Vorstandsvorsitzende und Geschäftsführerin des Deutschen Instituts für Terminologie (DIT), Fachbeirätin des Deutschen Terminologie-Tags (DTT), Vorsitzende des DIN-AA 01 „Grundlagen der Terminologiearbeit", stellvertretende Vorsitzende des DIN-NAT-Beirats und Mitglied im Rat für Deutschsprachige Terminologie (RaDT) der UNESCO.

Kontaktadresse
petra.drewer@h-ka.de
www.h-ka.de

IDS: Wörterbuch der Konnektoren

Angelika Ottmann

Im Rahmen seiner Online-Dienste bietet das Leibniz-Institut für Deutsche Sprache (IDS) ein Wörterbuch der Konnektoren an.

Es enthält grammatisch relevante Informationen, Beispiele und Belege zu Syntax und Semantik von Konnektoren wie *aber*, *weil*, *wohlgemerkt*, *sogar* oder *geschweige denn*, basierend auf den Ergebnissen der beiden Teile des Handbuchs der deutschen Konnektoren (https://www.ids-mannheim.de/gra/abgeschlosseneprojekte/konnektoren/). Die Suche ist auf Wortbasis oder auf Formularbasis möglich.

Der **wortbasierte Zugang** erfolgt mit Hilfe einer alphabetischen Liste (siehe Abb. 1). Über diese Liste lassen sich Informationen zu den einzelnen Konnektoren abrufen. Grammatische Fragen zu Konnektoren, beispielsweise zur Syntax oder zur Wortstellung in Nebensätzen, lassen sich somit auf einfache Weise klären.

Ein Eintrag enthält Informationen zu syntaktischer Klasse, Stellung im Satz und Bedeutung des betreffenden Konnektors. Zum Konnektor „weil“ beispielsweise wird die syntaktische Klasse als „Subjunktor“ mit Beispielen untermauert, es werden Beispiele für verschiedene Positionen im Satz angeführt, die semantische Klasse „kausal“ wird angegeben und schließlich wird noch die Bedeutung von „weil“ erläutert (siehe Abb. 2 auf Seite 37).

Bei der **formularbasierten Suche** (Auswahllisten am linken Rand) lassen sich mit Hilfe wortübergreifender Kategorien („Syntaktische Klasse“, „Semantische Klasse“, „Stellung“ und „Stil“) verschiedene Filter setzen und kombinieren (siehe Abb. 3 auf Seite 37).

Ein Beispiel: Bei Filterung nach der semantischen Klasse „kausal“ wird die Konnektorenliste in Abb. 4 auf Seite 37 angezeigt.

Das Wörterbuch der Konnektoren ist ein interessantes Nachschlagewerk vor allem für Spracharbeitende in den Bereichen Übersetzung, Redaktion, Lektorat usw. Es kann über folgenden Link aufgerufen werden:

https://grammis.ids-mannheim.de/konnektoren

Schauen Sie mal rein!

Angelika Ottmann
ottmann@dttev.org

Quelle:
Leibniz-Institut für Deutsche Sprache: „Wörterbuch der Konnektoren“. Grammatisches Informationssystem grammis. DOI: 10.14618/wb-konnektoren | https://grammis.ids-mannheim.de/konnektoren

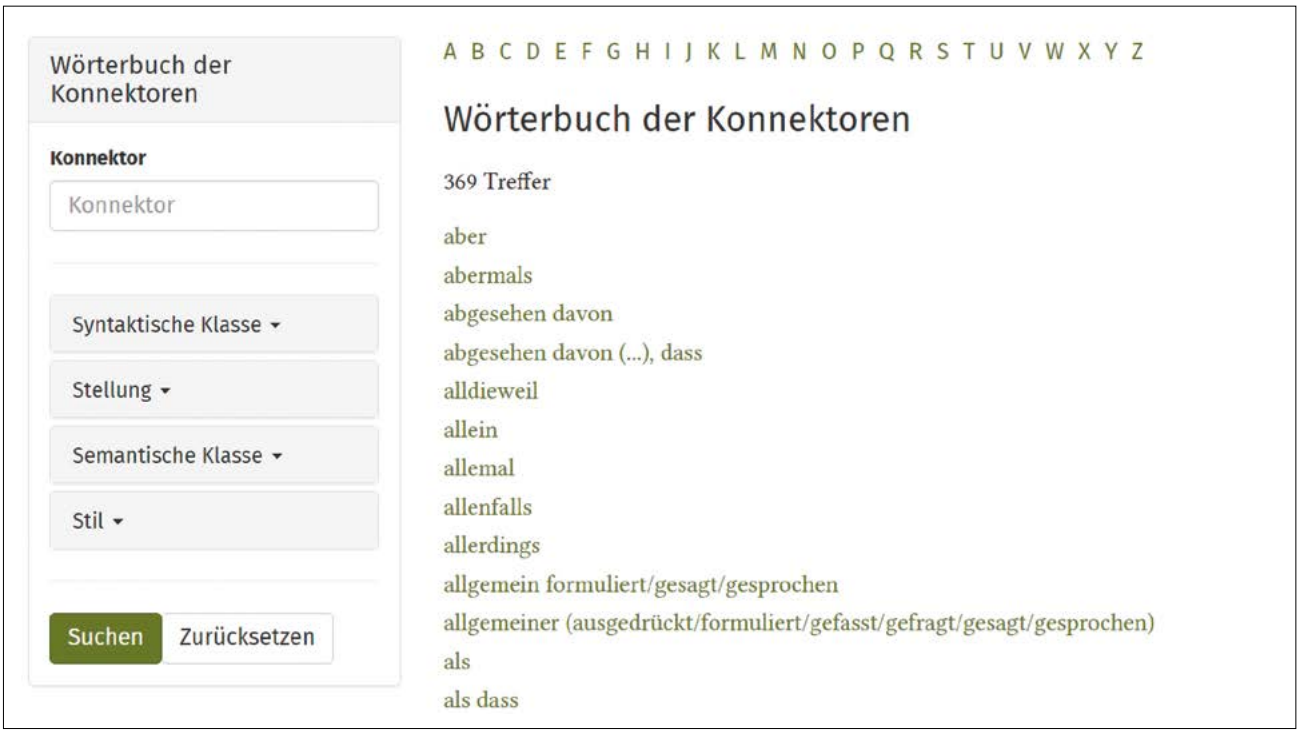


Abb. 1: Konnektorsuche – alphabetische Liste (Quelle: eigener Screenshot)

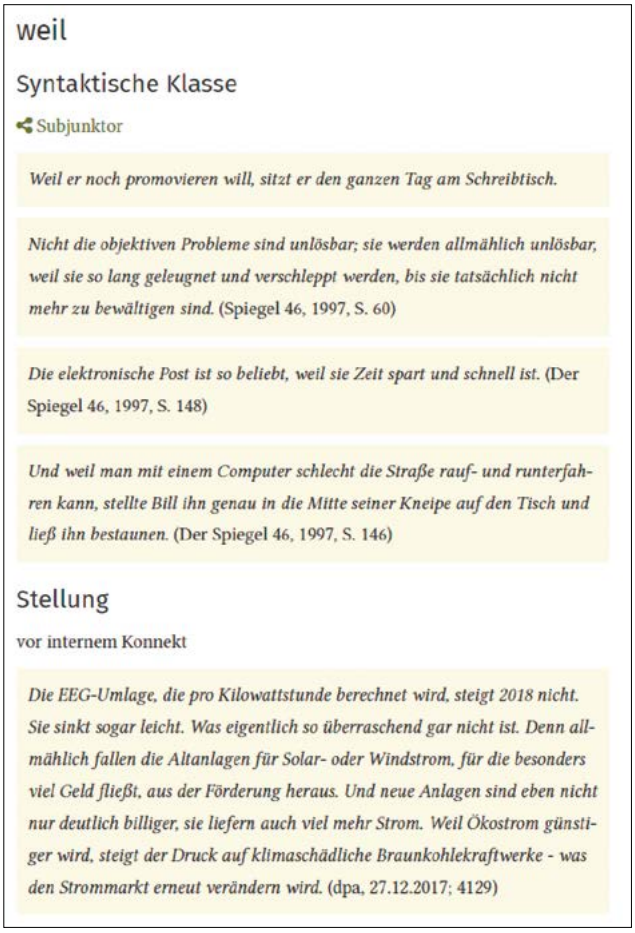


Abb. 2: Teileintrag zu Konnektor „weil“ (Quelle: eigener Screenshot)

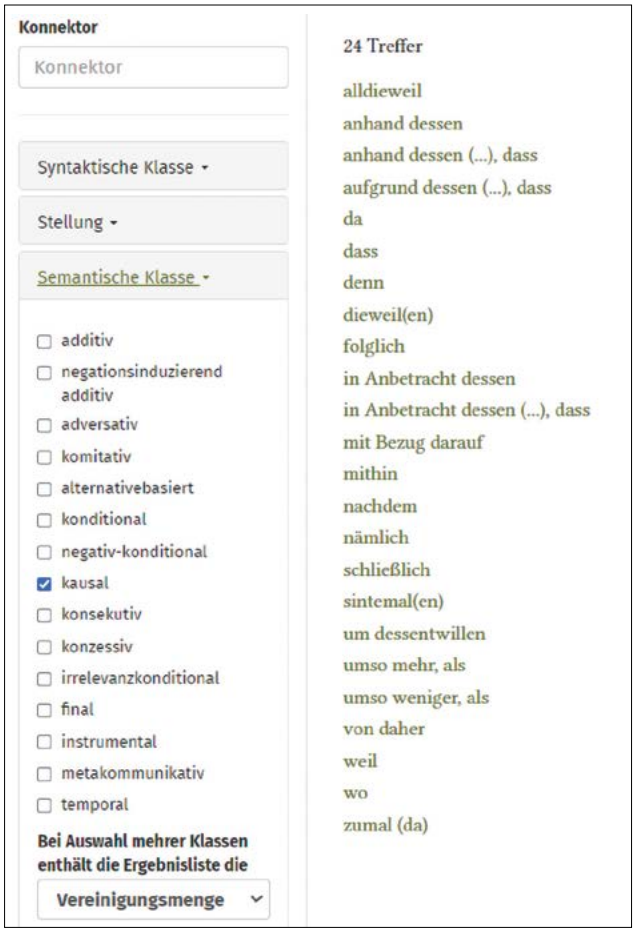


Abb. 4: Suche nach kausalen Konnektoren (Quelle: eigener Screenshot)

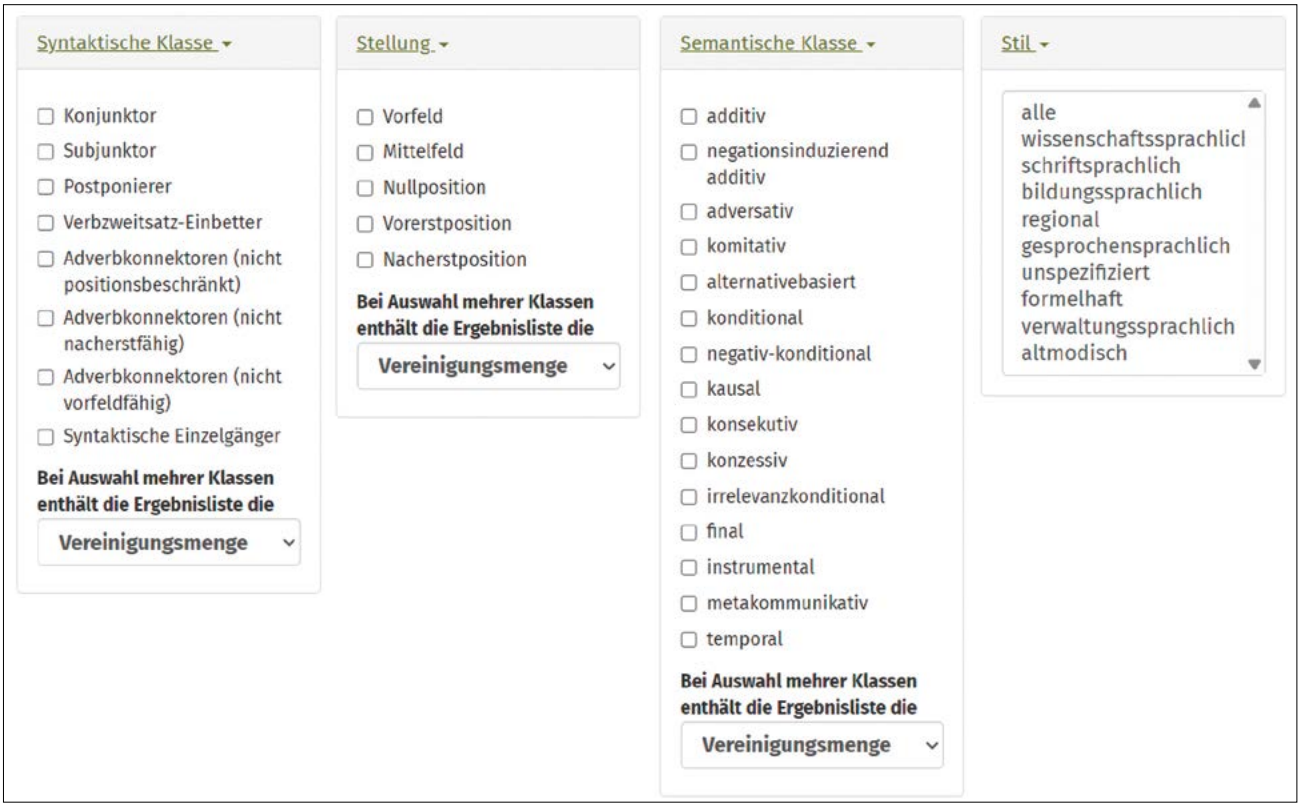


Abb. 3: Formularbasierte Suche mit Hilfe wortübergreifender Kategorien (Quelle: eigener Screenshot)

NFDI4ING-Konferenz 2025

Tom Winter

Vom 15. bis 16. September fand in Darmstadt die NFDI4ING-Konferenz 2025 statt. Sie ist Teil der Nationalen Forschungsdateninfrastruktur (NFDI), einer bundesweiten Initiative, die darauf abzielt, Forschungsdaten in Deutschland systematisch zu erschließen, zu vernetzen und nachhaltig verfügbar zu machen. NFDI4ING adressiert dabei die Ingenieurwissenschaften und entwickelt Standards, Werkzeuge und Konzepte für den Umgang mit wissenschaftlichen Daten entlang der FAIR-Prinzipien (Findable, Accessible, Interoperable, Re-usable).

Die Konferenz brachte Teilnehmer aus Wissenschaft, Forschung und Wirtschaft zusammen, um über Strategien und technische Entwicklungen für eine nachhaltige, interoperable Forschungsdatenlandschaft zu diskutieren. Das Programm umfasste Keynotes, Fachvorträge, Workshops und Diskussionsrunden zu Themen wie FAIR Digital Objects, Ontologien, Digitale Zwillinge und semantische Datenvernetzung.

Abb. 1: Die Konferenz fand bei schönstem Wetter im Georg-Christoph-Lichtenberg-Haus der TU Darmstadt statt. Blick in den Garten. (Foto: Institut für Fluidsystemtechnik (FST), TU Darmstadt)

Terminologiearbeit als Fundament für Ontologien und Wissensmodellierung

Ein zentrales Thema war die Bedeutung der Terminologiearbeit als semantische Grundlage für Ontologien und Wissensmodelle. In mehreren Beiträgen wurde deutlich, dass ohne präzise definierte und abgestimmte Begriffe keine konsistenten Ontologien entstehen können. Fehlende terminologische Klarheit führt zu semantischen Inkonsistenzen und Missverständnissen – insbesondere in interdisziplinären Projekten.

Ontologien gewinnen erst dann an Aussagekraft, wenn sie auf einer stabilen terminologischen Basis beruhen. Diese schafft die Voraussetzungen, um Konzepte, Relationen und Attribute in Datenmodellen eindeutig zu beschreiben und miteinander zu verknüpfen. Terminologiearbeit ist damit unverzichtbar für die Wissensmodellierung und entsprechend als fester Bestandteil von Daten- und Forschungsinfrastrukturen zu betrachten.

Ontologien als Dokumentationsgrundlage in Wissenschaft und Technik

Ein weiterer Schwerpunkt lag auf der Rolle von Ontologien als Dokumentations- und Modellierungsinstrument in wissenschaftlichen Disziplinen. Sie ermöglichen eine formale und nachvollziehbare Beschreibung komplexer Systeme und Prozesse, etwa in Ingenieurwissenschaften, Umweltforschung oder Mobilität.

Ontologien dienen der maschinellen Verarbeitung und der strukturierten Dokumentation von Wissen. Sie fördern die Interoperabilität unterschiedlicher Datenmodelle und schaffen ein gemeinsames Verständnis über Fachgrenzen hinweg. Mehrere Beiträge zeigten, dass Ontologien zunehmend als nachhaltige Dokumentationsgrundlage und als verbindendes Element zwischen Forschung, Anwendung und Standardisierung eingesetzt werden.

FAIR Digital Objects und FAIR Data Points

Ein weiteres Leitthema war der Übergang zu FAIR Digital Objects (FDOs) – also Datenobjekten, die selbst FAIR-Prinzipien erfüllen und damit adressierbar, interoperabel und wiederverwendbar sind. In diesem Zusammenhang wurden

FAIR Data Points (FDPs) als technische Schnittstellen vorgestellt, über die solche Objekte auffindbar und zugänglich gemacht werden.

Ziel ist eine Dateninfrastruktur, in der Objekte über standardisierte Metadaten, persistente Identifikatoren und klar definierte Zugriffsrechte verfügen. Ontologien und kontrollierte Vokabulare spielen dabei eine zentrale Rolle, da sie die semantische Beschreibung dieser Objekte unterstützen und den Informationsaustausch zwischen Systemen erleichtern.

Digitale Zwillinge auf dem Vormarsch

Und dann war da natürlich das Zukunftsthema „Digitaler Zwilling“ – also digitale Abbilder physischer Systeme, die kontinuierlich mit aktuellen Daten versorgt werden. Sie dienen der Simulation, Zustandsüberwachung und vorausschauenden Steuerung technischer Anlagen.

Diskutiert wurde, wie sich Digitale Zwillinge mit Ontologien und FAIR Digital Objects kombinieren lassen, um semantisch angereicherte, interoperable Modelle zu schaffen. Industriebeiträge aus Fertigung, Stadtplanung und Logistik verdeutlichten den Nutzen Digitaler Zwillinge: Sie ermöglichen es, Produkte, Prozesse und Entscheidungen durch Echtzeitanalyse, Simulation und prädiktive Steuerung effizienter, kostengünstiger und nachhaltiger zu gestalten.

Voraussetzung dafür: konsistente Wissensmodelle.

Fazit

Die NFDI4ING-Konferenz zeigte eindrucksvoll, dass semantische Interoperabilität zur Schlüsselfunktion moderner Forschungs- und Dateninfrastrukturen wird. Für das Feld der Terminologie bedeutet das:

- Terminologiearbeit bleibt zentral, um semantische Konsistenz in Ontologien und Datenmodellen sicherzustellen.
- Ontologien entwickeln sich zu universellen Dokumentations- und Strukturierungsinstrumenten in Forschung und Technik.
- FAIR Digital Objects und FAIR Data Points bilden die Basis für eine modulare, vernetzte Datenarchitektur.
- Digitale Zwillinge gewinnen in Ingenieurwesen und Industrie an Bedeutung und erfordern eine saubere semantische Modellierung.
- Der kontinuierliche Dialog zwischen Forschung und Praxis ist Voraussetzung für tragfähige Standards und nachhaltige Datenökosysteme.

Terminologiearbeit wird damit zum semantischen Fundament der digitalen Wertschöpfung – und zu einem entscheidenden Erfolgsfaktor für die Integration von Wissensmanagement, Künstlicher Intelligenz und datengetriebenen Prozessen. Ran ans Werk!

Weitere Informationen zu NFDI4ING finden Sie unter https://nfdi4ing.de/

Tom Winter
winter@dttev.org

Abb. 2: Organisationsteam und Teilnehmende der NFDI4ING-Konferenz 2025 in Darmstadt (Foto: Institut für Fluidsystemtechnik (FST), TU Darmstadt)

KALEIDOSCOPE

Unleash the Full Potential of Terminology

Enterprise-ready
Modular
Innovativ

QUICKTERM

Die führende Terminologieplattform für maximale Produktivität.

www.kaleidoscope.at