

## eTranslation – das maschinelle Übersetzungssystem der Europäischen Kommission

Werner Schweibenz<sup>1</sup> – Bibliotheksservice-Zentrum Baden-Württemberg (BSZ), Konstanz

### Mehrsprachigkeit und maschinelle Übersetzung im Kulturbereich

Die Mehrsprachigkeit von Informationsangeboten nimmt im Kultursektor eine immer wichtigere Rolle ein. Für Kultureinrichtungen wie Bibliotheken, Archive, Museen und Sammlungen ist es wichtig, ihrem Publikum Informationen nicht nur auf Deutsch, sondern auch in anderen Sprachen zu übermitteln, beispielsweise in Form von Ausstellungstexten und -katalogen, Audioguides oder auf Webseiten. Allerdings sind professionelle Übersetzungen arbeits-, kosten- und verwaltungsaufwendig, sodass sie im Tagesgeschäft für kurzfristig benötigte Texte oder für Übersetzungen mit dem Charakter von Überblicksinformationen häufig nicht extern vergeben werden können. Aus diesen Gründen ruhen große Hoffnungen auf dem Einsatz von Systemen für die maschinelle Übersetzung, also die automatische Übersetzung von Texten aus einer Sprache in eine andere durch ein Computerprogramm (vgl. Wikipedia-Eintrag „Maschinelle Übersetzung“). Ein Beispiel ist das Projekt *Europeana Translate* (1. Mai 2021–30. April 2023), in dem es um die Übersetzung der Metadaten von mehr als 25 Millionen Datensätzen der *Europeana* ging. Bekannte Anbieter auf diesem Gebiet sind beispielsweise *Google Translate*, *Microsoft Translator* oder *DeepL Translate*, um nur einige zu nennen. Weniger bekannt ist aber, dass auch die Europäische Kommission ein maschinelles Übersetzungssystem für ein breites Publikum anbietet: *eTranslation*.

### eTranslation – Maschinelles Übersetzungssystem der Europäischen Kommission<sup>2</sup>

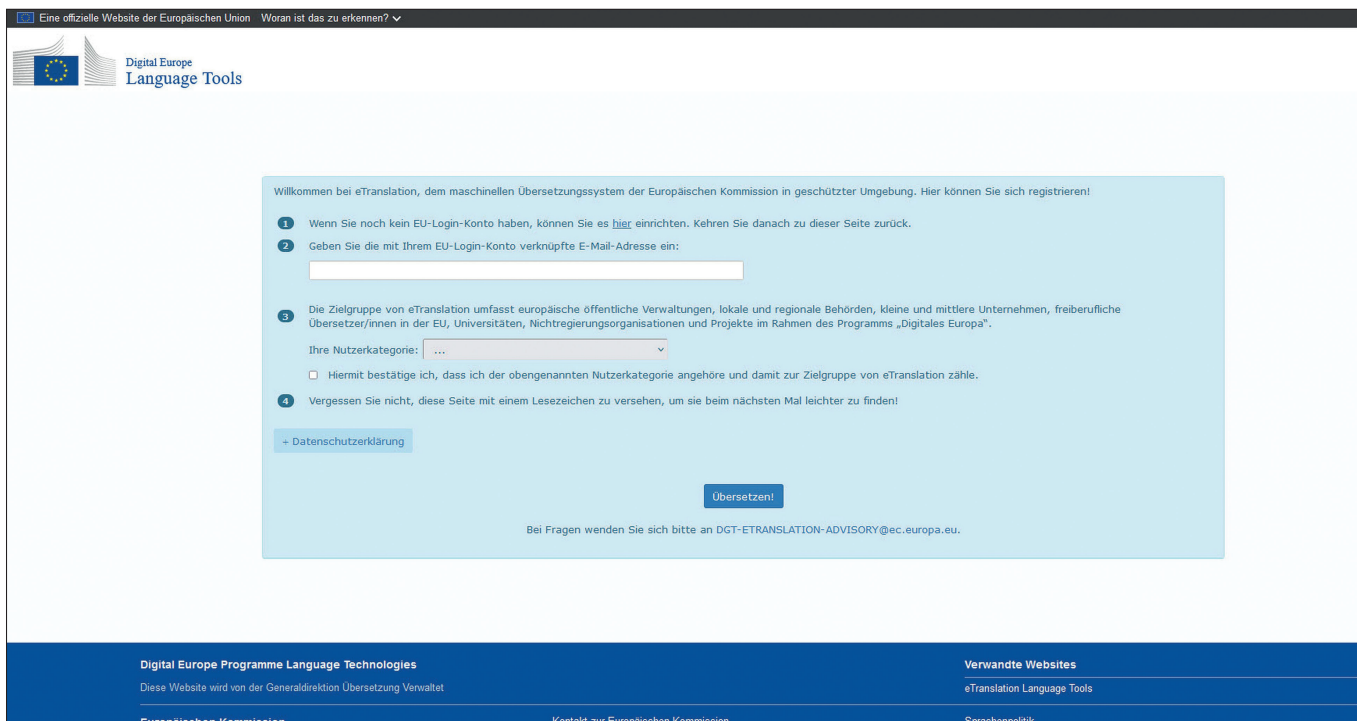
*eTranslation* ist das maschinelle Übersetzungssystem der Europäischen Kommission, das am 15. November 2017 offiziell eingeführt wurde und aktuell in der Version 12.3 vom 31. März 2023 angeboten wird. Anfangs diente es ausschließlich für die Übersetzung von internen Dokumenten der europäischen Institutionen in die derzeit 24 Amts- und Arbeitssprachen der Europäischen Union. Im Rahmen des Arbeitsprogramms *Digitales Europa* (2021–2027) zur Bereitstellung digitaler Technologien für Unternehmen, Bürger und öffentliche Verwaltungen wird das System einem breiten Nutzerkreis zugänglich gemacht. So können öffentliche Verwaltungen und Einrichtungen, Universitäten,

freiberufliche Übersetzer\*innen, kleine und mittlere Unternehmen (KMU) sowie Nichtregierungsorganisationen *eTranslation* kostenlos nutzen; erforderlich ist lediglich eine Registrierung, in deren Rahmen man eine Nutzerkennung erhält.

Das System ermöglicht Übersetzungen aus bzw. in alle EU-Amtssprachen, also Bulgarisch, Dänisch, Deutsch, Englisch, Estnisch, Finnisch, Französisch, Griechisch, Irisch, Italienisch, Kroatisch, Lettisch, Litauisch, Maltesisch, Niederländisch, Polnisch, Portugiesisch, Rumänisch, Schwedisch, Slowakisch, Slowenisch, Spanisch, Tschechisch und Ungarisch. Hinzu kommen folgende Sprachen: Arabisch, Chinesisch, Isländisch, Japanisch, Norwegisch, Russisch, Türkisch und Ukrainisch. Bei der Nutzung ist zu beachten, dass die Europäische Kommission keine Haftung für mögliche Fehler oder Unzulänglichkeiten der Übersetzung übernimmt.

*eTranslation* ist nach eigenen Angaben ein neuronales maschinelles Übersetzungssystem. Vereinfacht gesagt arbeitet es mit Algorithmen, die mit übersetzten Textpaaren darauf trainiert wurden, statistische oder musterbasierte Analysen von Texten durchzuführen und so automatisch zu lernen, wie neue Texte zu übersetzen sind. Dafür werden sehr große Mengen an übersetzten Textpaaren benötigt, bei denen die Texte eine Übersetzung aus der ersten Sprache (Ausgangssprache) in die zweite Sprache (Zielsprache) sind (sogenannte Parallelkorpora). Dazu werden Daten in einem Online-Repository gesammelt: *ELRC-share* (siehe Abschnitt Training von *eTranslation* mit Textpaaren aus dem Kulturbereich).

Das System bietet eine hohe Sicherheit in Bezug auf die Dokumente, die zur Verarbeitung hochgeladen werden, wobei es in der Verantwortung der Nutzer\*innen liegt, ob sie möglicherweise vertrauliche Dokumente an *eTranslation* übermitteln wollen bzw. dürfen. Die hochgeladenen Texte werden nur im Datenverarbeitungszentrum der Europäischen Kommission gespeichert und verarbeitet (in einer Cloud von *Microsoft Azure*); sie werden nicht mit Dritten geteilt. Hochgeladene Texte verbleiben 24 Stunden auf den Servern und werden dann automatisch gelöscht. Der Löschmodus kann von den Nutzer\*innen mit der Verarbeitungsoption NACH DEM HERUNTERLADEN LÖSCHEN (DELETE AFTER DOWNLOAD) beschleunigt werden; mit dieser Option werden die hochgeladenen Texte sofort



**Abb. 1**  
Registrierungsmaske von eTranslation, Zugriffsdatum: Juni 2023.

nach der Rücklieferung der Übersetzungen an die Nutzer\*innen gelöscht.

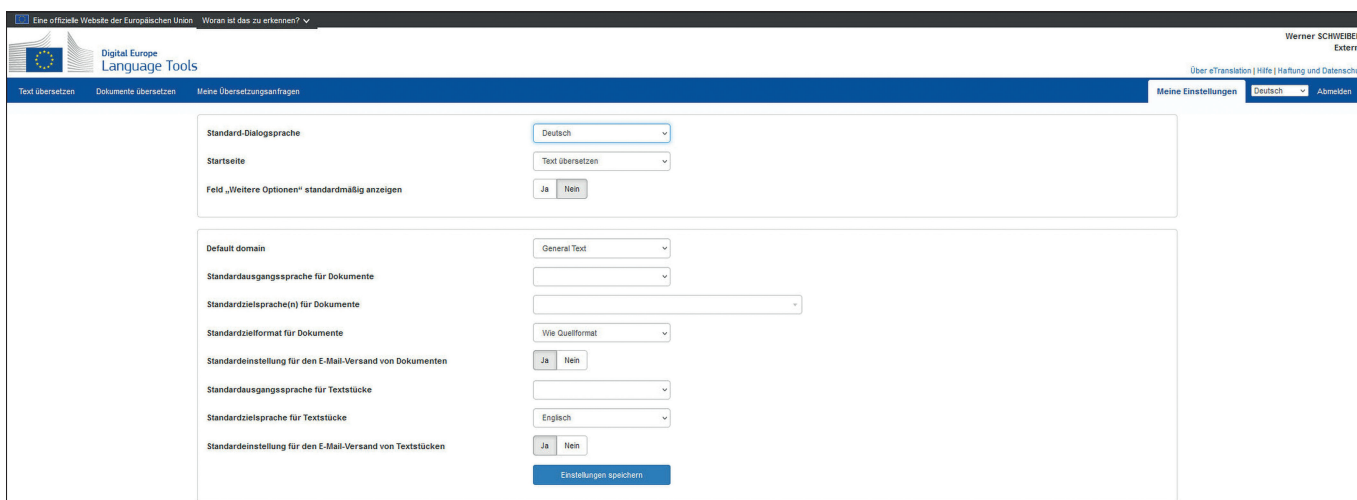
Ebenso transparent wie beim Umgang mit den Texten verhält sich das System in Bezug auf die Speicherung personenbezogener Daten. So erfasst eTranslation bestimmte Informationen, um die Nutzung des Dienstes und die Bearbeitung von Anfragen zu ermöglichen sowie zu statistischen Zwecken. Dazu gehören die Nutzerkennung, die Zugangszeiten, die gewählten Sprachen der Übersetzungen, die Größe des zu übersetzenden Textes und die Domäne der E-Mail-Adresse der Nutzenden (Nutzername@ec.europa.eu). Diese Daten werden 18 Monate aufbewahrt und dann archiviert. Die persönliche E-Mail-Adresse, an die die Übersetzungen zugestellt werden, wird nur für den Versand genutzt, nicht aber archiviert. Die Europäische Kommission gibt die erhobenen Nutzer-

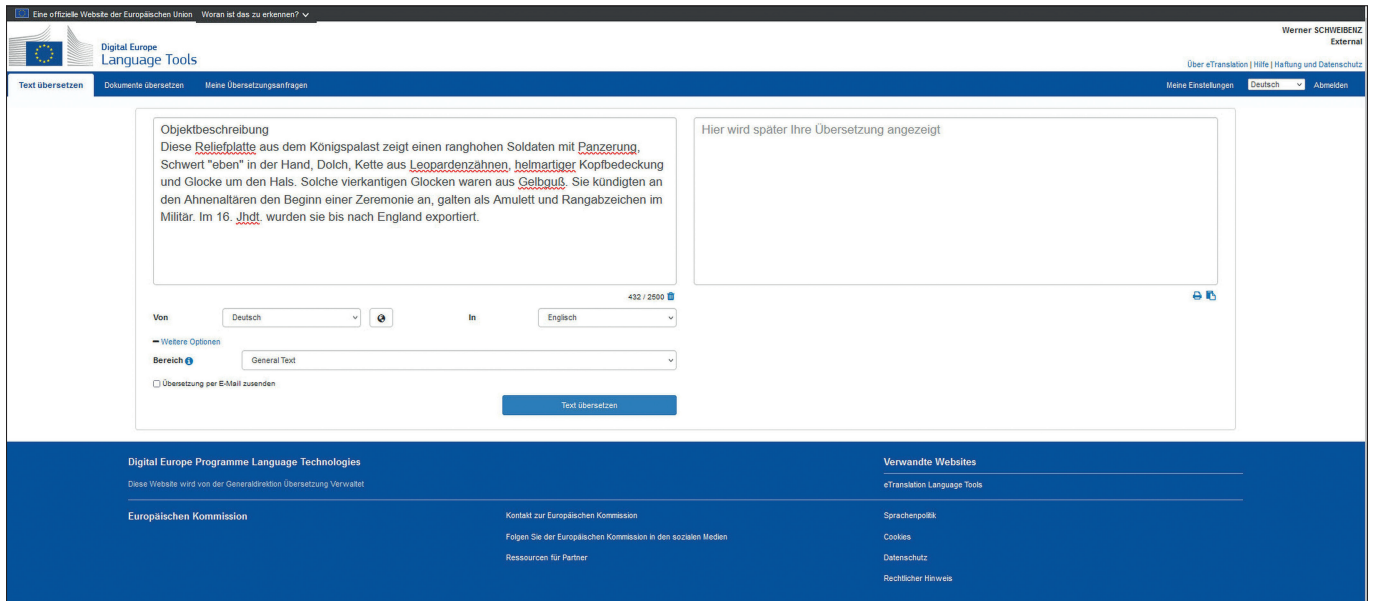
daten nicht an Dritte weiter, sofern sie dazu nicht gesetzlich verpflichtet ist.

### Registrierung für die Nutzung von eTranslation

Für die Nutzung von eTranslation ist eine Registrierung mit einer persönlichen, in der Regel dienstlichen E-Mail-Adresse erforderlich. Bei der Anmeldung wird ein EU-Login-Konto erzeugt (Nutzername@ec.europa.eu), das mit der E-Mail-Adresse verknüpft wird. Die einzelnen Schritte sind auf der Registrierungsseite angegeben. In Ermangelung einer passgenaueren Kategorie kann man unter Punkt 3 der Registrierung bei der Nutzerkategorie ÖFFENTLICHE VERWALTUNG angeben. Für Fragen ist unten auf der Seite eine Kontaktadresse beim eTranslation-Team angegeben.

**Abb. 2**  
Meine Einstellungen in eTranslation, Zugriffsdatum: Juni 2023.





### Einstellungen in eTranslation

eTranslation bietet die Möglichkeit, das System für den persönlichen Gebrauch zu konfigurieren. Dazu dient die Registerkarte MEINE EINSTELLUNGEN rechts oben, die auf eine Konfigurationsseite führt, auf der persönliche Präferenzen festgelegt werden können, wie zum Beispiel die STANDARD-DIALOGSPRACHE, die STARTSEITE, den Übersetzungsstil – hier als DEFAULT DOMAIN bezeichnet – sowie sprachbezogene Angaben zu den Dokumenten (Abb. 2).

### Benutzung von eTranslation

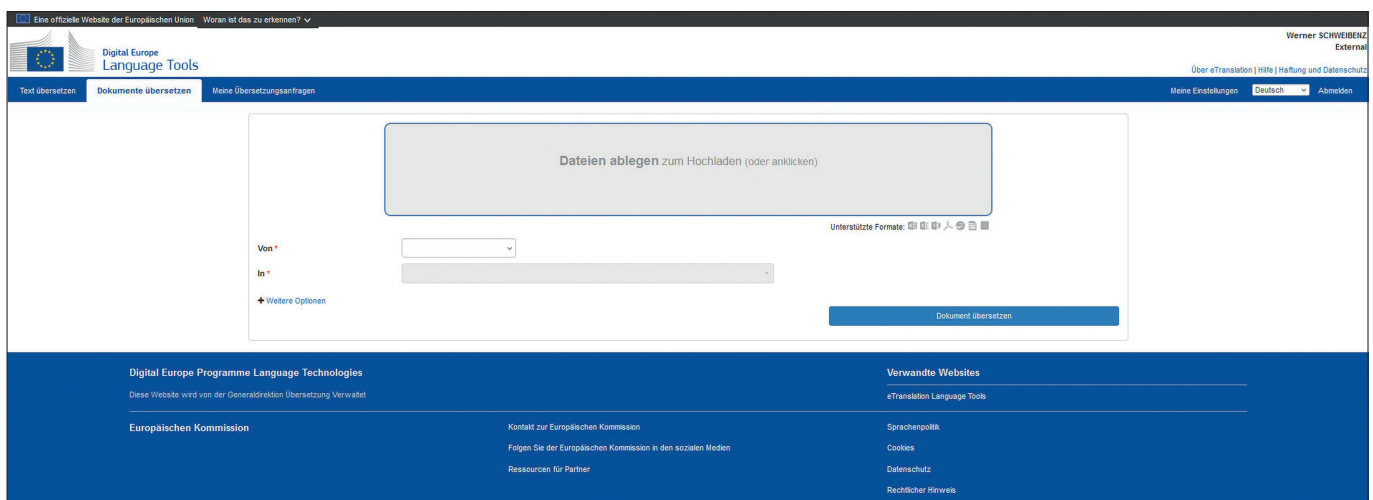
Bei der Benutzung von eTranslation kann man zwischen zwei Angaben der Text- bzw. Dokumentart wählen, die durch die Registerkarten TEXT ÜBERSETZEN und DOKUMENTE ÜBERSETZEN repräsentiert werden. Im Folgenden wird die Bedienung erläutert. Dabei ist zu beachten, dass man pro Tag höchstens 100 Übersetzungen mit eTranslation durchführen darf. Als Übersetzung zählt auch,

wenn dasselbe Dokument in mehrere Sprachen übersetzt wird.

Die Registerkarte TEXT ÜBERSETZEN (Abb. 3) ist für kürzere Texte gedacht und nimmt die Verarbeitung im Browser-Fenster vor. Dieses besteht aus zwei Fenstern: In das linke Fenster für die Ausgangssprache wird der zu übersetzende Text ohne Formatierungen eingetragen, die Länge beträgt maximal 2.500 Zeichen. Dann werden in den Feldern darunter die Angaben zur Übersetzung ausgewählt. Im Feld VON wird die Ausgangssprache angegeben. Die Option SPRACHE ERKENNEN (AUTOMATISCHE SPRACHERKENNUNG, LANGUAGE DETECTION), also die automatische Erkennung der Ausgangssprache, sollte nur dann verwendet werden, wenn die Sprache tatsächlich unbekannt ist, weil sich aufgrund der aufwendigen Analyse der Sprache die Antwortzeit deutlich erhöht. Im Feld IN wird die Zielsprache ausgewählt. Die Auswahl WEITERE OPTIONEN öffnet ein Menü, in dem der BEREICH des Übersetzungsstils ausgewählt werden kann (siehe Abschnitt Rohübersetzungen mit ver-

**Abb. 3** Registerkarte TEXT ÜBERSETZEN, Zugriffsdatum: Juni 2023.

**Abb. 4** Registerkarte DOKUMENTE ÜBERSETZEN, Zugriffsdatum: Juni 2023.



schiedenen Übersetzungsstilen). Mit der Schaltfläche TEXT ÜBERSETZEN wird der Übersetzungsvorgang gestartet. Das Ergebnis wird im rechten Fenster angezeigt. Leider fehlt eine Sanduhr oder ein Fortschrittsbalken, sodass man ohne Rückmeldung geduldig die Anzeige der Übersetzung abwarten muss. Unter dem Fenster der Zielsprache kann man durch die Vergabe von Sternen die Qualität der Übersetzung bewerten.

Die Registerkarte DOKUMENTE ÜBERSETZEN (Abb. 4) ist für die Verarbeitung von längeren Texten gedacht, die hier hochgeladen werden; die unterstützten Dateiformate werden unter dem Fenster als kontextsensitive Icons (mit Mouse-Over) angezeigt. Die Einstellungen zur Übersetzung werden vorgenommen wie oben beschrieben, zusätzlich kann noch ein Ausgabeformat für das übersetzte Dokument gewählt werden. Die Übersetzung kann auch unter der Registerkarte MEINE ÜBERSETZUNGSANFRAGEN heruntergeladen werden. Das Dokument steht 24 Stunden zum Herunterladen zur Verfügung.

Sowohl bei der Übersetzung im Browser als auch mittels Dokument besteht die Möglichkeit, bestimmte Textteile von der Übersetzung auszuschließen. Dies passiert über Textauszeichnungen in Form von sogenannten Tags. Mit dem Tag [notranslate] wird der Beginn der Aussparung gekennzeichnet, mit dem Tag [/notranslate] das Ende. Dabei ist zu beachten, dass in *Microsoft Word* die beiden Tags identisch formatiert sein müssen, weil sie sonst nicht erkannt werden und die Übersetzung nach dem ersten Tag abbricht.

*eTranslation* bietet eine Schnittstelle (API) für Maschine-zu-Maschine-Interaktionen von zugelassenen Parteien an. Einzelheiten zu Funktionsweise und Registrierung finden sich auf der Hilfeseite (nur für registrierte Nutzer).

Im Gegensatz zu einigen Produkten von kommerziellen Anbietern hat *eTranslation* keinen sogenannten Translation Memory. Das ist ein Übersetzungsarchiv, in dem zusammengehörige Textsegmente getrennt nach Ausgangs- und Zielsprache gespeichert werden. Mit einer solchen Funktion können bestimmte, häufig wiederkehrende Wörter oder Phrasen stets in derselben, einmal kontrollierten und festgelegten Weise übersetzt werden.

### Rohübersetzungen mit verschiedenen Übersetzungsstilen

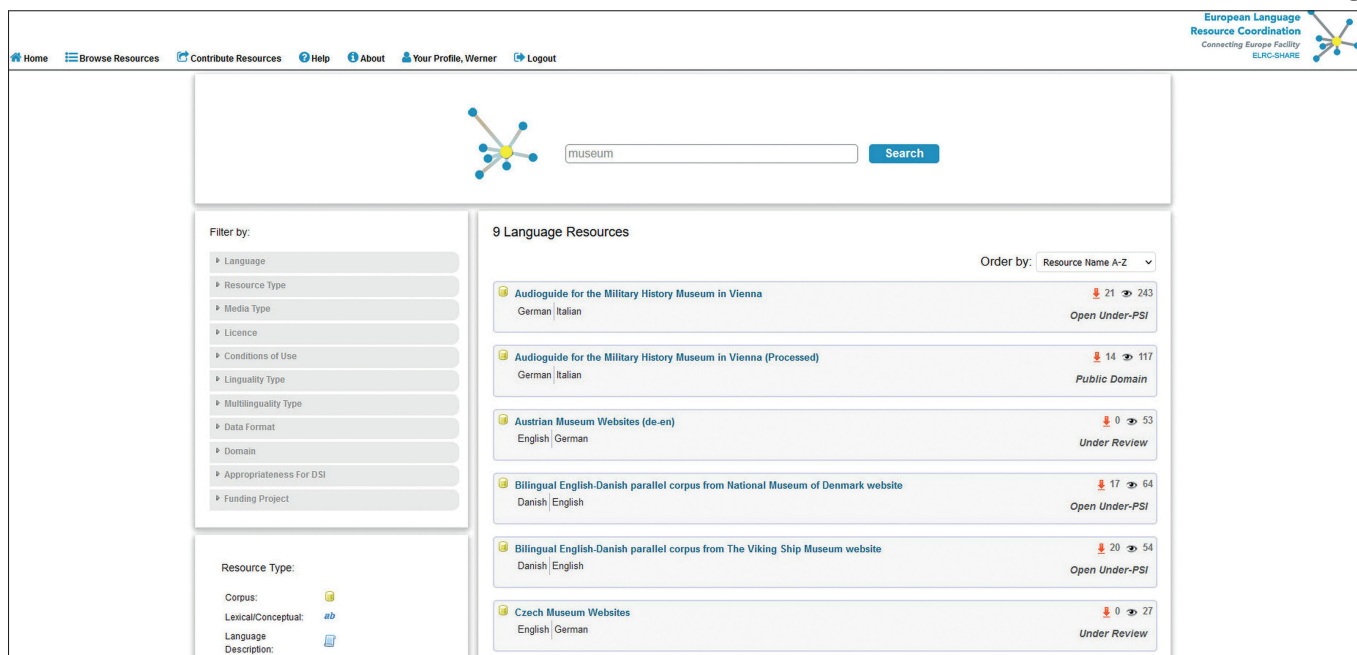
*eTranslation* liefert maschinelle Rohübersetzungen, die man zum Erfassen des Textinhalts oder als Ausgangspunkt für eine menschliche Übersetzung bzw. Überarbeitung nutzen kann. Das heißt, dass diese Rohfassungen noch professionell überarbeitet werden müssen, um präzise, qualitativ hochwer-

tige Übersetzungen zu erhalten. Denn maschinelle Übersetzungen sind nicht so zuverlässig wie menschliche Übersetzungen, aber dafür billiger und schneller (Foti 2022, S. 34). Dass ein Bedarf an solchen maschinellen (Roh-)Übersetzungen gegeben ist, zeigt die beachtliche Zahl der Seiten, die im Jahr 2019 mit *eTranslation* übersetzt wurden: 96 Millionen (Foti 2022, S. 33).

Zu beachten ist, dass die Qualität der maschinellen Übersetzung stark schwanken kann – je nach gewähltem Sprachpaar (z. B. Deutsch-Englisch als Sprachpaar mit umfangreichem Textkorpus, Deutsch-Arabisch als Sprachpaar mit weniger umfangreichem Textkorpus), Texttyp (z. B. Rechtstext, Verwaltungstext, Ausstellungstext) und Übersetzungsstil. Unter einem Übersetzungsstil, auch als Bereich oder Domäne bezeichnet, versteht man einen Filter, der bei der Übersetzung Faktoren wie Sprachstil, Ausdruck und Vokabular steuert. Für das „Erlernen“ des entsprechenden Übersetzungsstils wurde und wird das maschinelle Übersetzungssystem mit qualitativ hochwertig übersetzten Textpaaren trainiert (siehe Abschnitt Training von *eTranslation* mit Textpaaren aus dem Kulturbereich). Auf der Basis von sehr großen Textpaaren kann das Übersetzungssystem aufgrund von Analysen mit hoher Wahrscheinlichkeit voraussagen, welche Wörter in der Folge zum bisherigen Text passen und diese entsprechend auswählen. Aktuell bietet *eTranslation* mehrere Stile an, von denen „General Text“ und „EU Formal Language“ die am häufigsten genutzten sind (Foti 2022, S. 33 f.). „General Text“ eignet sich gut für allgemeinsprachliche Dokumente, also Gebrauchstexte im weitesten Sinne, während „EU Formal Language“ auf dem Übersetzungsarchiv der Europäischen Institutionen basiert und besonders für formelle, juristische und Verwaltungstexte geeignet ist. Der Übersetzungsstil „Europeana (GLAMs – Galleries, Libraries, Archives, Museums)“ (vormals „Culture“) für den Kulturbereich ist im Aufbau.

### Training von *eTranslation* mit Textpaaren aus dem Kulturbereich

Seit Sommer 2021 wird eine Sammlung von Textdokumenten aus dem Kulturbereich angelegt, um damit das maschinelle Übersetzungssystem für einen Übersetzungsstil „Europeana“ zu trainieren. Dafür werden in der *European Language Resource Coordination (ELRC)* zwei- oder mehrsprachige Textpaare gesucht. *ELRC* wurde bereits 2015 ins Leben gerufen, um Sprachressourcen zur Verbesserung der Qualität, Leistungsfähigkeit und Sprachenabdeckung von *eTranslation* zu sammeln und bereitzustellen. Dies geschieht in einem Online-Repository namens *ELRC-share*, das beim *Deutschen Forschungszentrum für Künstliche Intelligenz*



European Language Resource Coordination  
Connecting Europe Facility  
ELRC-SHARE

Home Browse Resources Contribute Resources Help About Your Profile, Werner Logout

Search: museum

Filter by:

- Language
- Resource Type
- Media Type
- Licence
- Conditions of Use
- Linguality Type
- Multilinguality Type
- Data Format
- Domain
- Appropriateness For DSI
- Funding Project

Resource Type:

- Corpus:
- Lexical/Conceptual:
- Language:
- Description:

9 Language Resources

Order by: Resource Name A-Z

Resource Name	Language(s)	Downloads	Views	Status
Audioguide for the Military History Museum in Vienna	German Italian	21	243	Open Under-PSI
Audioguide for the Military History Museum in Vienna (Processed)	German Italian	14	117	Public Domain
Austrian Museum Websites (de-en)	English German	0	53	Under Review
Bilingual English-Danish parallel corpus from National Museum of Denmark website	Danish English	17	64	Open Under-PSI
Bilingual English-Danish parallel corpus from The Viking Ship Museum website	Danish English	20	54	Open Under-PSI
Czech Museum Websites	English German	0	27	Under Review

(DFKI) Saarbrücken angesiedelt ist. Dieses Repository wird vom *Common European Language Data Space* abgelöst, der am 19. Januar 2023 gegründet wurde.

In vielen Bereichen und einigen Sprachen stellt die Verfügbarkeit von qualitativ hochwertigen Sprachpaaren eine Herausforderung dar (ELRC White Paper 2022, S. 24), jedoch nicht im Kulturbereich im deutschsprachigen Raum. Hier liegen bereits sehr große Mengen von hochwertig übersetzten Sprachpaaren vor, beispielsweise in Form von Ausstellungstexten und -katalogen, als Audioguides oder auf Webseiten. Ein wichtiger Aspekt ist die Frage, wie man die Verfasser\*innen/Urheber\*innen dieser Dokumente überzeugen kann, sie *ELRC-share* zur Verfügung zu stellen (ELRC White Paper 2022, S. 24). Wenn diese Bereitschaft gegeben wäre, bliebe noch das Problem der personellen und finanziellen Ressourcen, um die Dokumente der Sprachpaare in den Kulturinstitutionen einzusammeln, online an *ELRC-share* zu übermitteln und dort für das Training von *eTranslation* aufzubereiten. Diese Probleme sind bekannt und lösbar (ELRC White Paper 2022, S. 26f.); sie könnten zeitnah im Rahmen eines EU-Projekts in Angriff genommen und gelöst werden. Dabei wäre eine aktive Unterstützung durch die *Europeana*, die *Deutsche Digitale Bibliothek* und die *Beauftragte der Bundesregierung für Kultur und Medien* sicherlich sehr hilfreich.

## Zusammenfassung und Ausblick

*eTranslation* bietet ein kostenloses Übersetzungssystem, das dem neusten Stand der Technik entspricht und die europäischen Standards in Bezug auf Datenschutz und -sicherheit erfüllt. Bei kommerziellen Anbietern muss man prüfen, ob ihre Dienstleistungsangebote den gesetzlichen Anforderungen entsprechen.

Das System hat sich im Rahmen des Arbeitsprogramms *Digitales Europa* zur Bereitstellung digitaler Technologien deutlich weiterentwickelt und wird dies aufgrund der aktuellen Förderung voraussichtlich auch in den nächsten Jahren tun. Mit dem Übersetzungsstil „General Text“ lassen sich bereits jetzt Gebrauchstexte aus dem Kulturbereich in ansprechender Qualität als Rohfassungen übersetzen. Erste praktische Erfahrungen werden gerade im Rahmen des *Netzwerks Koloniale Kontexte* gesammelt. Es handelt sich um eine Initiative von ethnologischen Museen, Sammlungen und Bibliotheken zu Fragen der digitalen Zusammenführung, Bearbeitung und Sichtbarkeit sowie Nutzung von Materialien und Daten aus kolonialen Kontexten. Hier rückte die Mehrsprachigkeit von Informationen im Rahmen der Kommunikation mit den Herkunftsgesellschaften in den Mittelpunkt, sodass ein großer Bedarf an schnell und kostengünstig verfügbaren (Roh-)Übersetzungen besteht. Durch ein kontinuierliches Training des Übersetzungsstils „Europeana“ könnte mittelfristig eine deutliche Verbesserung der Übersetzungsqualität entsprechend den Anforderungen des Kultursektors erreicht werden.

## Abb. 5

ELRC-share Sprachressourcen „Museum“, Zugriffsdatum: Juni 2023.

### Literatur und Quellenangaben

[letzter Zugriff jeweils: 14.07.2023]

Amts- und Arbeitssprachen der Europäischen Union:  
[https://european-union.europa.eu/principles-countries-history/languages\\_de](https://european-union.europa.eu/principles-countries-history/languages_de)

European Language Resource Coordination (2022), ELRC White Paper. AI for a Multilingual Europe. Why Language Data Matters, third print edition, Saarbrücken 2022, online unter <https://www.clarin.gr/sites/default/files/ELRC-White-Paper.pdf>.

eTranslation – Maschinelles Übersetzungssystem der Europäischen Kommission:  
[https://commission.europa.eu/resources-partners/etranslation\\_de](https://commission.europa.eu/resources-partners/etranslation_de)

eTranslation – Anmeldeseite:  
<https://webgate.ec.europa.eu/etranslation/public/welcome.html>

eTranslation – Hilfe:  
<https://webgate.ec.europa.eu/etranslation/help.html>

EU-Arbeitsprogramm Digitales Europa:  
<https://digital-strategy.ec.europa.eu/de/activities/work-programmes-digital>

Europeana Translate:

<https://pro.europeana.eu/project/europeana-translate>

Foti, Markus (2022), *eTranslation: Le système de traduction automatique de la Commission européenne en appui d'une Europe numérique*. In: Traduire – Revue française de la traduction (2022), 246, S. 28–35, online siehe <https://journals.openedition.org/traduire/2725>.

Maschinelle Übersetzung – Wikipedia:  
[https://de.wikipedia.org/wiki/Maschinelle\\_%C3%9Cbersetzung](https://de.wikipedia.org/wiki/Maschinelle_%C3%9Cbersetzung)

Netzwerk Koloniale Kontexte:  
<https://www.evifa.de/delueber-uns/fid-projekte/netzwerk-koloniale-kontexte>

- 
1. Der Autor lernte *eTranslation* im Rahmen seiner Entsendung als *Nationaler Sachverständiger zur beruflichen Weiterbildung* zur EU-Kommission nach Brüssel (10/2017–03/2018) kennen und nutzt es seitdem regelmäßig.
  2. Alle Angaben zu *eTranslation* in diesem Beitrag stammen aus dem Web-Angebot der Europäischen Kommission.