



Maschinelle Übersetzung

Whitepaper der D.O.G. GmbH / Autor: Dr. François Massion

Maschinelle Übersetzung

Eine Alternative, die sich lohnt

In diesen schwierigen Zeiten sind Unternehmen mehr denn je auf der Suche nach Möglichkeiten, Kosten zu sparen. Gleichzeitig wollen sie so schnell wie möglich zusätzlichen Umsatz generieren. Kunden im Ausland sollen dazu beitragen. Mehr übersetzen, schneller übersetzen und weniger bezahlen? Es klingt wie die Quadratur des Kreises.

Fast jeder hat inzwischen mit automatischen Übersetzungsprogrammen wie Google Translate oder DeepL experimentiert. Die Ergebnisse sind beeindruckend. Zwar stimmt nicht alles und diese Systeme machen auch grobe Fehler. Aber folgende Frage ist berechtigt: Kann nicht zumindest ein Teil des Übersetzungsbedarfs maschinell erstellt werden?

Jeder, der ernsthaft über den Einsatz maschineller Übersetzung (MÜ) nachdenkt, sieht sich vielen Unbekannten gegenüber.

- ✓ **Wie funktioniert MÜ?**
- ✓ **Gibt es dafür unterschiedliche Programme?**
- ✓ **Wie ist die Qualität?**
- ✓ **Wieviel kann man damit sparen?**

Ziel dieses White Papers ist es, einige Antworten und Empfehlungen zu liefern.

Inhalt

Kurzer historischer Rückblick	2
Wie maschinelles Übersetzen heute funktioniert	2
Encoder und Decoder	3
Systemauswahl: Generisch oder Individuell?	4
Welche Fehler machen MÜ-Systeme?	5
Datensicherheit	6
Post-Editieren	6
Korrekturtechnologie – ErrorSpy.....	7
Was können Sie tun, um die Ergebnisse zu optimieren?.....	9
Terminologie für Maschinen	10
Besseres Trainingsmaterial	11
Kosten- und Zeitvorteile	11
Wofür ist MÜ geeignet?	12



Kurzer historischer Rückblick

MÜ hat eigentlich eine lange Geschichte. Zu Beginn des Kalten Krieges war es für die Großmächte und v.a. für die Vereinigten Staaten und die damalige Sowjetunion wichtig, schnell zu verstehen, worüber der Gegner schreibt.

Der erste Durchbruch gelang im Januar 1954. Ein IBM 701-Großrechner übersetzte 60 russische Sätze mit der „legendären“ Geschwindigkeit von 2,5 Zeilen pro Sekunde ins Englische.

Seit diesem Ereignis sind über 65 Jahre verstrichen. In der Zwischenzeit hat es viele Höhen und Tiefen gegeben. Der anfänglichen Begeisterung folgte eine Phase der Ernüchterung. Unterschiedliche Technologien wurden der Reihe nach eingesetzt: regelbasiertes Übersetzen dann beispielbasierte Verfahren in den 60er Jahren. In den 90er Jahren leitete statistisches maschinelles Übersetzen (SMT – Statistical Machine Translation) erneut eine Wende ein. SMT ermittelt statistisch aus zahlreichen Paralleltexten, welche Wörter oder Ausdrücke in der

Ausgangs- und Zielsprache zusammen vorkommen. Die Übersetzungsergebnisse sind trotz holprigem Stil relativ gut. Heute wird SMT immer noch mit Erfolg eingesetzt.

Schließlich leitete Google Translate 2016 die Ära des neuronalen maschinellen Übersetzens (NMT Neural Machine Translation) ein. NMT-Systeme benutzen Verfahren der Künstlichen Intelligenz (KI) und zeichnen sich durch ihren sehr flüssigen Stil aus. Die Ergebnisse sind teilweise überraschend gut, so gut sogar, dass sie heute manchmal „Humanparität“ erreichen, d.h. einige Übersetzungen sind so gut wie menschliche Übersetzungen.

Wie maschinelles Übersetzen heute funktioniert

Um zu beurteilen, was Sie von MÜ erwarten und welche Fehler auftreten können, sollten Sie zunächst verstehen, wie neuronales MÜ funktioniert.

Maschinelle Übersetzungssysteme benutzen Modelle, die mit neuronalen Netzen trainiert werden. Für jedes Sprachpaar trainieren Machine-Learning-Spezialisten

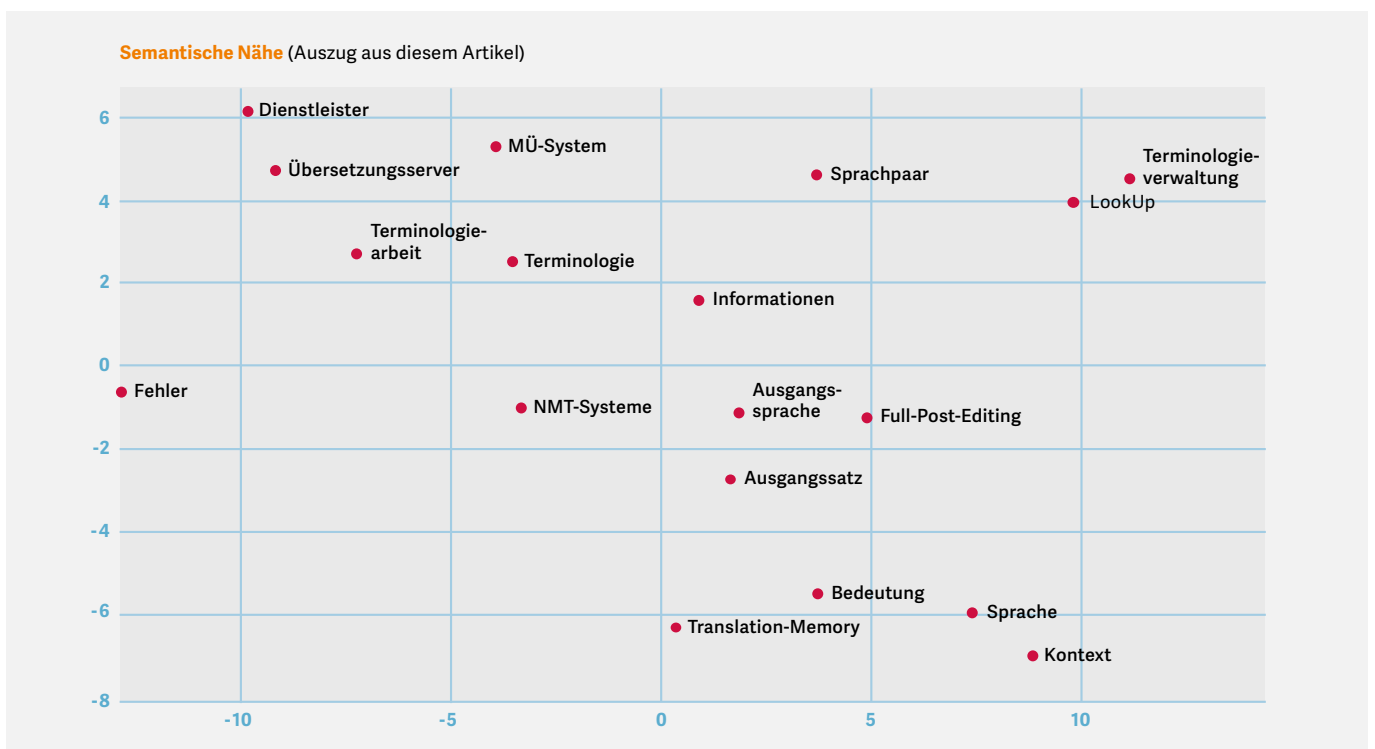


Abbildung 1: Gelernte Semantik



ein Modell. Ein Übersetzungsprogramm benutzt ein gelerntes Modell, um neue unbekannte Texte zu übersetzen. Die Sprachkombination Deutsch-Englisch verwendet ein anderes Modell als die Kombination Englisch-Deutsch.

Was das Modell erreichen kann, lässt sich vielleicht am besten am Beispiel des Spracherwerbs von Kindern veranschaulichen: Wenn ein kleines Kind seine Sprache lernt, merkt es sich Wörter und Wortfolgen, die im Laufe der Jahre immer wieder gemeinsam verwendet werden. Je öfter Wörter zusammen erscheinen, desto besser erinnert sich das Kind an sie: **Ich bin müde. Ich möchte schlafen** ist für das Kind normal und gehört zu seinem gelernten „Modell“, während **Ich bin müde. Ich möchte das Flugzeug** zwar bekannte Wörter sind, die aber nicht zu seinem „Modell“ passen.

Encoder und Decoder

Ein neuronales maschinelles Übersetzungssystem benutzt zwei verknüpfte Netzwerke: den Encoder und den Decoder.

Als Input dienen riesige Sammlungen von übersetzten Sätzen: jeweils ein Ausgangssatz mit seiner Übersetzung. Die Übersetzung ist sozusagen die Zielvorgabe, die das System fehlerfrei lernen muss.

Der Encoder liest die Ausgangssätze ein und lernt welche Wörter einen gemeinsamen Kontext haben. Da eine Maschine nur Zahlen versteht, wandelt sie die Bedeutung von Wörtern und Sätzen in Zahlenvektoren um. Am Ende des Lernprozesses liefert der Encoder für jeden Ausgangssatz einen Satzvektor, der die Bedeutung des Satzes in Zahlen wiedergibt.

In einem zweiten Schritt lernt die Maschine wie dieser Vektor in eine Fremdsprache übertragen wird. Diesen Schritt steuert der Decoder mit dem Ausgangssatz als Input und der Übersetzung als Output. In mehreren Zwischenschritten lernt der Decoder Merkmale, die für die Übersetzung von Wörtern der Ausgangssprache relevant sind.

Das gelernte Modell kann nun neue, bisher unbekannte Sätze übersetzen. Der Algorithmus berechnet für jeden Satz die Übersetzung, die die höchste Wahrscheinlichkeit hat. Da die Sprache nicht immer präzise und oft mehrdeutig ist, ist die vorgeschlagene Übersetzung manchmal falsch bzw. sie kann stilistisch weniger gelungen sein als ähnliche korrekte Übersetzungen, die beim Algorithmus weniger gepunktet haben.

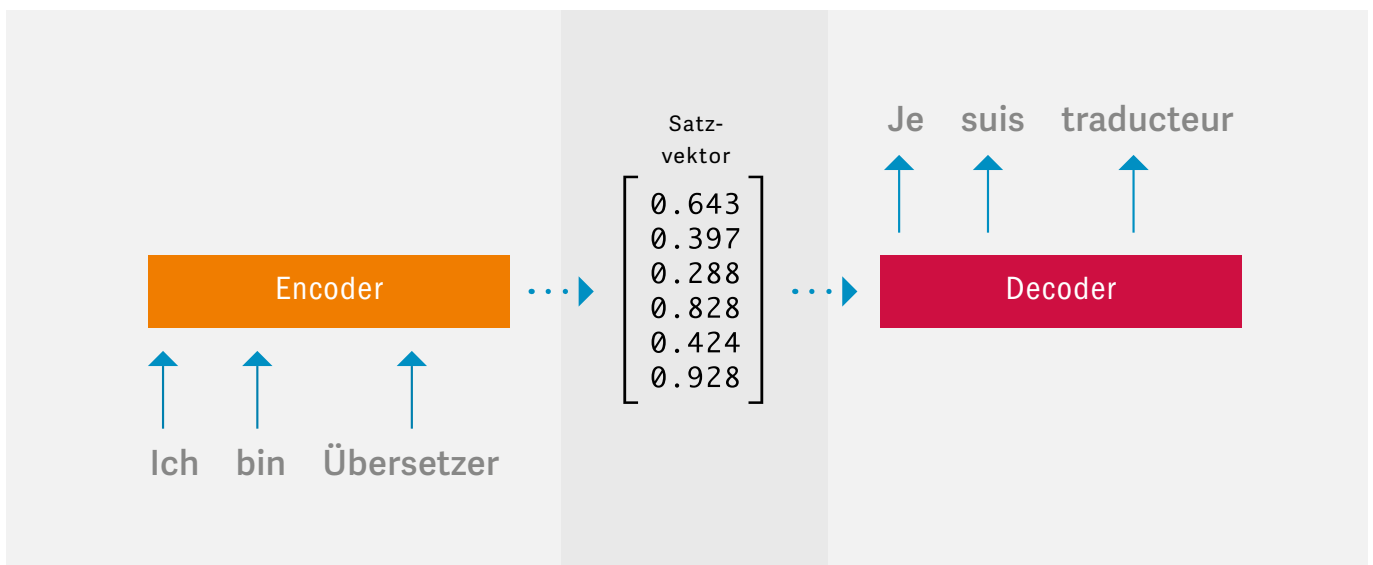


Abbildung 2: Encoder und Decoder



Systemauswahl: generisch oder individuell?

Wenn Sie also maschinell übersetzen möchten, stehen Ihnen mehrere Optionen zur Verfügung. Die Auswahl ist nicht einfach, denn MÜ ist nicht gleich MÜ. Je mehr Sie maschinell übersetzen möchten, desto mehr benötigen Sie eine individuelle Lösung. Sie werden Systeme brauchen, die mit Ihrer Firmensprache trainiert wurden und Ihre Fachterminologie beherrschen.

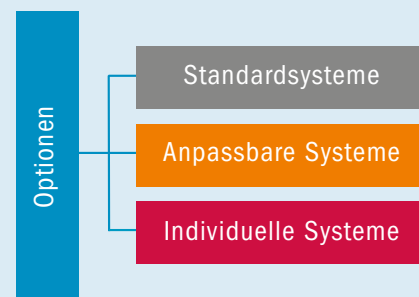
Das Herzstück eines maschinellen Übersetzungssystems ist ein neuronales Netz. Dies haben alle Systeme gemeinsam. Ein wesentlicher Anteil der heute eingesetzten Systeme verwendet Open Source Bibliotheken wie OpenNMT (<https://opennmt.net/>), die sie konfigurieren und mit eigenen Skripten und Programmen erweitern.

Je mehr Sie sich in Richtung individuelles MÜ-System bewegen, desto mehr entsprechen die erzeugten Übersetzungen dem Sprachgebrauch Ihres Unternehmens. Dadurch fällt weniger Korrekturaufwand an.

Das ist z. B. der Fall, wenn ein Unternehmen bestimmte Termini (**Batteriefachdeckel** statt **Akkuschutzdeckel**) oder Ausdrucksweisen (**Öffnen Sie den Batteriefachdeckel** statt **Batteriefachdeckel öffnen**) bevorzugt. Da außerdem viele Wörter abhängig vom Kontext unterschiedliche Bedeutungen haben können, liefern öffentlich zugängliche Übersetzungssysteme oft die am meisten verbreitete Bedeutung. Sie liefern nicht unbedingt die spezifische Übersetzung, die das Unternehmen verwendet (das sog. „Long Tail“-Phänomen). Wörter wie **Leistung**, **Gerät**, **Anlage** oder **Scheibe** haben mehr als eine „richtige“ Übersetzung. Ferner bietet eine maßgeschneiderte Lösung eine größere Sicherheit bezüglich des Datenschutzes.

Drei Optionen

Je nachdem wie generisch oder firmenspezifisch die Übersetzungsergebnisse sein sollen, stehen Unternehmen drei Ansätze zur Auswahl:



- 1.** Sie können direkt oder indirekt über API öffentlich zugängliche Übersetzungstools wie Google Translate, DeepL oder Microsoft Translator verwenden. Diese Systeme sind sofort einsetzbar und bereits konfiguriert. Sie berücksichtigen jedoch keine firmenspezifischen Themen und Terminologien.
- 2.** Sie können Plattformen wie Systran, Kantan oder Globalese verwenden. Sie erlauben in begrenztem Maße das Training und die Konfiguration eines MÜ-Systems.
- 3.** Sie entwickeln selbst bzw. lassen von einem spezialisierten Dienstleister ein maßgeschneidertes System entwickeln, das mit Texten und Informationsmaterial aus Ihrem Unternehmen trainiert wird.



Welche Fehler machen MÜ-Systeme?

Fehler von NMT-Systemen haben einige Gemeinsamkeiten:

Fehlertyp	Ausgangssatz	Übersetzung	Bemerkung
Inkonsistente oder nicht eingehaltene Terminologie	<ul style="list-style-type: none"> • Anzeige und Bedienelemente • Bedienelemente – Leiste • Auf der rechten Seite sind die Bedienelemente angeordnet. 	<ul style="list-style-type: none"> • Display and operating elements • Control elements – Bar • The controls are located on the right side. 	NMT-Systeme berechnen jede Übersetzung neu. Infolgedessen können sie sich nicht konsequent an eine vorgegebene Terminologie halten.
Missverständene Homonyme	NRW plant neue Corona-Auflagen	NRW plans new corona coatings	Kommt bei Oberbegriffen öfter vor.
Hinzufügungen	Qualifikation Personal	Required qualification of personnel	Bei ansonsten korrekten Übersetzungen nicht einfach zu entdecken.
Kontext-/Bezugfehler	It must be saved.	Es muss gerettet werden.	Sie (die Datei) muss gespeichert werden.

Wie eingangs erwähnt, produzieren NMT-Systeme teilweise verblüffend gute Übersetzungen. Wer einen Text maschinell übersetzt, wird eine Mischung zwischen diesen guten oder scheinbar guten Übersetzungen und falschen bzw. misslungenen Übersetzungen vorfinden.

Die genaue Art von Fehlern hängt von der Basistechnologie (SMT oder NMT) und vom gewählten Modell (generisch, anpassbar oder individuell) ab.

Warum tun sich NMT-Systeme so schwer mit Terminologie?

Eine der größten Herausforderungen neuronaler maschineller Übersetzungssysteme (NMT-Systeme) ist es, Terminologievorgaben zu berücksichtigen. Im Gegensatz zum Technologievorgänger (SMT, statistische maschinelle Übersetzungssysteme) ist es nach heutigem Stand der Technik quasi unmöglich, dem System eine Liste mit Terminologievorgaben zu geben.

Das liegt daran, dass NMT-Systeme komplexe neuronale Netze sind, die in unzähligen Operationen lernen, wie Wörter übersetzt werden. Sie speichern diese Informationen als Zahlenreihen in großen Vektoren, die die komplexen Beziehungen zwischen Wörtern beider Sprachen wiedergeben. Die Algorithmen können mit reinen Übersetzungstabellen (wie: **Scheibe = disk**) nichts anfangen.

siehe S. 6, Abb. 3: 'Bedeutung' eines Wortes für ein NMT-System →



```
In [22]: 1 print(model['übersetzen'])

[-0.3972198  0.15762967  0.46194917  0.17740306 -0.36603695  0.40951872
-0.18695748 -0.4747253  -0.37309977 -0.51228845  0.24443704 -0.15607826
-0.27267137 -0.25051364 -0.03269976 -0.23351745  0.41749892 -0.23172306
 0.41134125  0.33964443 -0.20515428  0.28367534  0.0628183  0.21196295
-0.03653887 -0.3595316  0.05570595  0.9420088  -0.01160865 -0.0248226
 0.08398762 -0.27247164  0.2707084  0.32289538 -0.11952855  0.05280286
-0.02814898  0.2553159  -0.7383995  0.4192747  0.30214256  0.23328382
-0.2331925  0.07106207 -0.38610226 -0.12182928 -0.06937908  0.20980522
 0.0566839  -0.24953169 -0.68693984  0.7137946  0.6530615  -0.40889168
 0.0336898  0.57632774  0.36043662  0.02917454  0.9916566  -0.16568148
-0.00262833  0.5754901  0.20918594 -0.10612661  0.01525148  0.02797612
 0.23608086  0.3928159  0.68412286  0.10793792  0.02760796 -0.34926915
 0.30073845  0.12777413 -0.16146977 -0.441489  -0.48704648 -0.3150954
-0.08529658 -0.3245209  -0.47360054  0.7129007  -0.02352894  0.04178019
 0.41401875 -0.30696347 -0.03826707 -0.41467956  0.48962304 -0.16006011
 0.28656143  0.39558902 -0.17324285  0.31181908  0.45250472 -0.50092775
 0.25884745 -0.24865492  0.35699815 -0.208062 ]
```

Abbildung 3: 'Bedeutung' eines Wortes für ein NMT-System

Datensicherheit

Viele der Informationen, die ein Unternehmen produziert, sind vertraulich: Spezifikationen, Angebote, Testberichte, Besprechungsnotizen usw. Wenn Mitarbeiter diese Informationen über öffentliche oder ungeschützte Übersetzungsserver automatisch übersetzen lassen, stellt dies ein hohes Risiko dar. Noch größer ist das Risiko, wenn sich die betreffenden Server außerhalb der Europäischen Union bzw. Deutschlands befinden.

Aus diesem Grund sind bei der Entscheidung über die Einführung eines Systems der Standort des Übersetzungsservers und die getroffenen Sicherheitsmaßnahmen sehr wichtig. Hier gibt es unterschiedliche Lösungen, die Sie im Hinblick auf Kosten und Sicherheit bewerten können:

- ✓ So können Sie beispielsweise den Übersetzungsserver in Ihren eigenen Räumlichkeiten hosten, was Ihnen einerseits ein hohes Maß an Sicherheit bietet, andererseits aber Zeit, Ressourcen und Kosten für die Bereitstellung und Wartung der Infrastruktur erfordert.

- ✓ Alternativ kann ein externer Dienstleister den Server bei sich bzw. bei einem der großen zuverlässigen etablierten Public-Cloud-Anbieter wie Amazon AWS, Google-Cloud-Plattform (GCP) oder Microsoft-Azure hosten.

Dadurch erreichen Sie ein hohes Maß an Sicherheit für Ihre Daten und Ihr geistiges Eigentum.

Post-Editieren

Nur in bestimmten Situationen ist eine rein maschinelle Übersetzung sinnvoll, z. B. wenn Sie Live-Chats übersetzen oder wenn große Informationsmengen sehr schnell verfügbar sein müssen. Auch hier sind mit Firmentexten trainierte Systeme besser geeignet.

Das gebräuchlichste Arbeitsmodell ist MÜ in Kombination mit Post-Editing. Den Output des maschinellen Übersetzungssystems korrigiert ein Mensch, der Post-Editor. Post-Editoren sind Fachleute, die über eine entsprechende Ausbildung verfügen, denn (1) Maschinen machen andere Fehler als Menschen und (2) je nach Qualitäts-



anspruch muss nicht alles korrigiert werden, was un- schön ist.

Es gibt inzwischen eine Norm für das Post-Editing (ISO 18587), die zwei Korrekturstufen vorsieht: eine leichte Nachbearbeitung (Light-Post-Editing), bei der hauptsächlich inhaltliche und Bedeutungsfehler korrigiert werden, und eine vollständige Nachbearbeitung (Full-Post-Editing), bei der die Übersetzung von ähnlicher Qualität sein muss wie die eines menschlichen Übersetzers.

Am Ende sind beim Full-Post-Editing die Ergebnisse genauso gut wie eine klassische Übersetzung. Wenn Sie sich davon ein Bild machen möchten, können Sie die englische Version unserer Website besuchen, die wir nach diesem Verfahren (MÜ + Post-Editing) erstellt haben (www.dog-gmbh.de/en). Wir konnten die Produktionskosten senken und vor allem konnten wir in diesem Fall die englische Version unserer Website viel schneller veröffentlichen.

Korrekturtechnologie - ErrorSpy

Wie prüft man effizient und zuverlässig größere Mengen an maschinell übersetzten Texten? Da Kosten und Zeit

Arbeit des Post-Editoren beim **light Post-Editing**

- ✓ **Ziel:** einen verständlichen Text erstellen und so wenig wie möglich ändern, aber:
 - ✓ Sinnfehler korrigieren
 - ✓ Missverständliche Übersetzungen umschreiben
 - ✓ Falsche Zahlen korrigieren
 - ✓ Sinnentstellende Schreibfehler korrigieren
- ✗ **NICHT:**
 - ✓ Stil verbessern (z. B. wörtliche Übersetzung)
 - ✓ Leichte Grammatikfehler korrigieren

Arbeit des Post-Editoren beim **full Post-Editing**

- ✓ **Ziel:** Übersetzung soll mit einer Humanübersetzung vergleichbar sein. Das bedeutet:
 - ✓ Sinnfehler korrigieren
 - ✓ Falsche Zahlen korrigieren
 - ✓ Rechtschreibfehler beheben
 - ✓ Stil und Grammatik optimieren
 - ✓ Terminologie vereinheitlichen

Relationsgraph für Begriff 1 (Leistung)

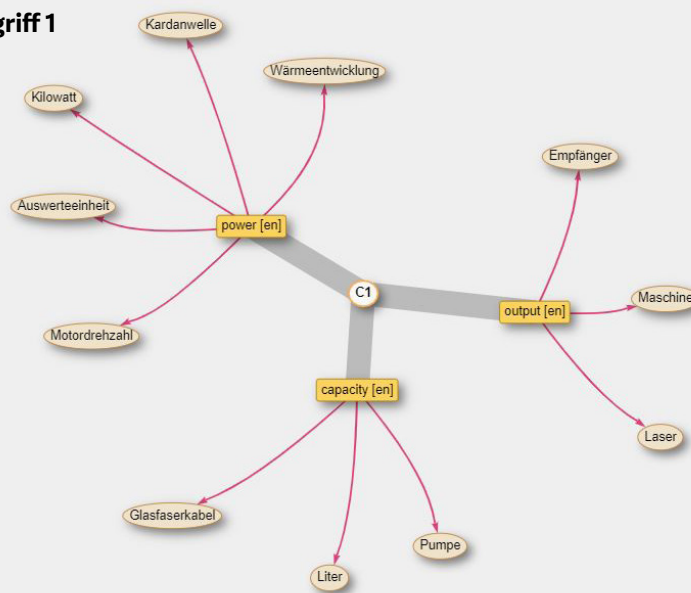


Abbildung 4: Kontextuelle Relationen in LookUp



wichtige Faktoren bei der Einführung von MÜ-Lösungen sind, ist der Post-Editor als Mensch in einer Zwickmühle. Er darf nicht zu langsam arbeiten, sonst verpuffen die Vorteile von MÜ + Post-Editing. Er darf aber auch nicht zu schnell arbeiten, sonst übersieht er wichtige Fehler. Dies umso mehr, wenn die maschinelle Übersetzung zuerst einmal gut klingt.

Was liegt also näher, als die Hilfe automatischer Prüfsysteme zur Unterstützung des Post-Editors in Anspruch zu nehmen? Schließlich ist es in der Industrie bereits Praxis: Optische Sensoren erkennen automatisch defekte Produkte am Fließband, Rechtschreibkorrektursysteme finden Schreibfehler in Dokumenten usw.

Seit bald 20 Jahren entwickelt und benutzt D.O.G. ErrorSpy, eine bewährte Qualitätssicherungssoftware, die Übersetzungen auf eine ganze Reihe von Fehlern (Zahlen, Terminologie, Vollständigkeit, Einheitlichkeit) prüft. ErrorSpy macht Post-Editoren auf mögliche Fehler aufmerksam.

Der leistungsfähige integrierte Editor ist eine große Hilfe bei der Fehlerkorrektur. Seit etwa 5 Jahren sind Prüffunktionen hinzugekommen, die speziell für maschinelle Übersetzungen Vorteile verschaffen. An vorderster Stelle steht die Kontextprüfung.

MÜ-Systeme tun sich besonders schwer, die richtige Übersetzung im Kontext zu verwenden. Wörter wie „Leistung“ haben beispielsweise sehr unterschiedliche Übersetzungen, je nachdem um welches Gerät es sich handelt. Ist es ein Motor, der eine Kraftleistung (EN: **power**) erzeugt? Ist es eine Pumpe, bei der es um eine Förderleistung in Liter (EN: **capacity**) geht? Oder ist es eine Maschine, die eine bestimmte Anzahl Schrauben pro Minuten herstellt (EN: **output**)? Dank der in der Terminologie hinterlegten Kontextrelationen kann ErrorSpy falsche MÜ-Übersetzungen im Kontext erkennen.

Eine genaue Beschreibung von ErrorSpy ist hier zu finden: www.dog-gmbh.de/produkte/errorspy

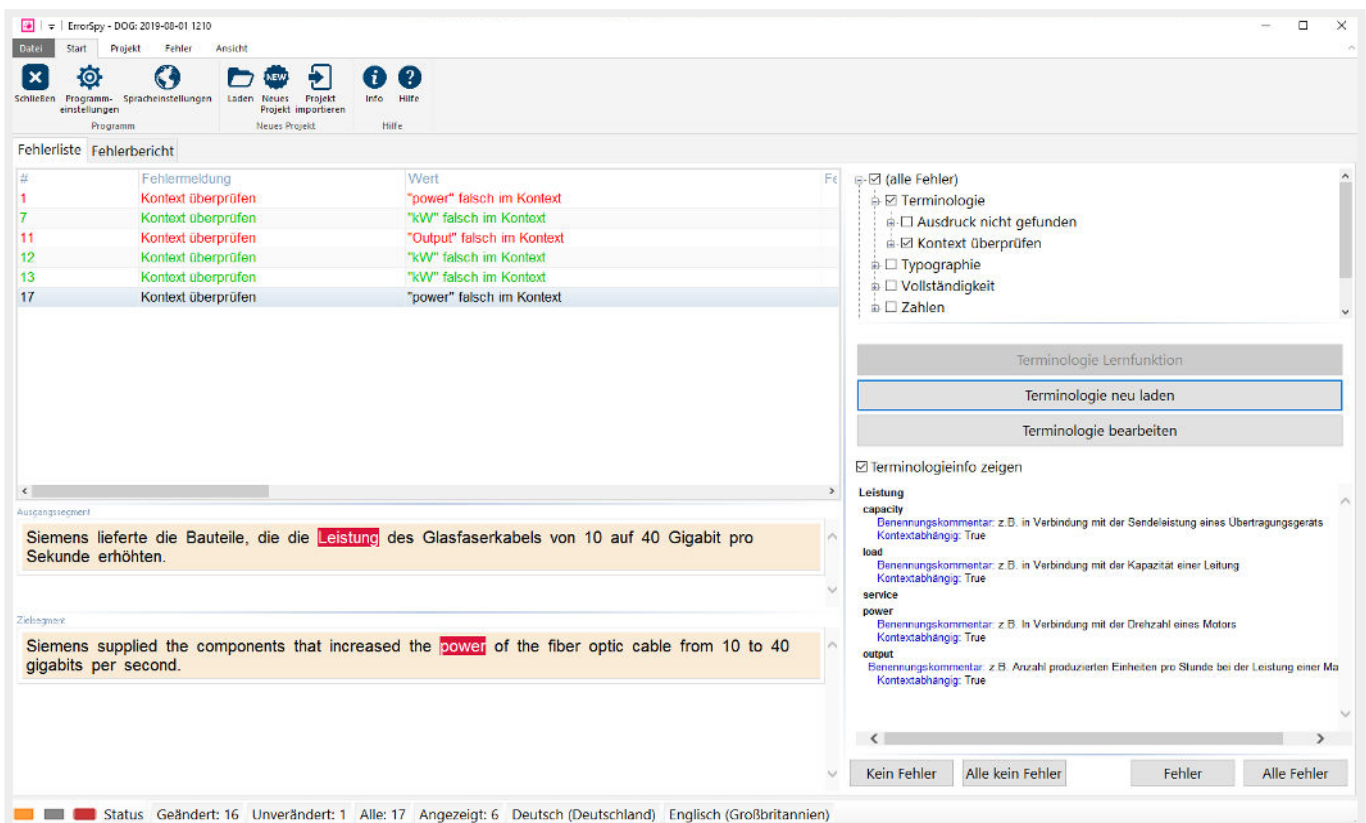


Abbildung 5: Einzigartige Kontextprüfung bei ErrorSpy



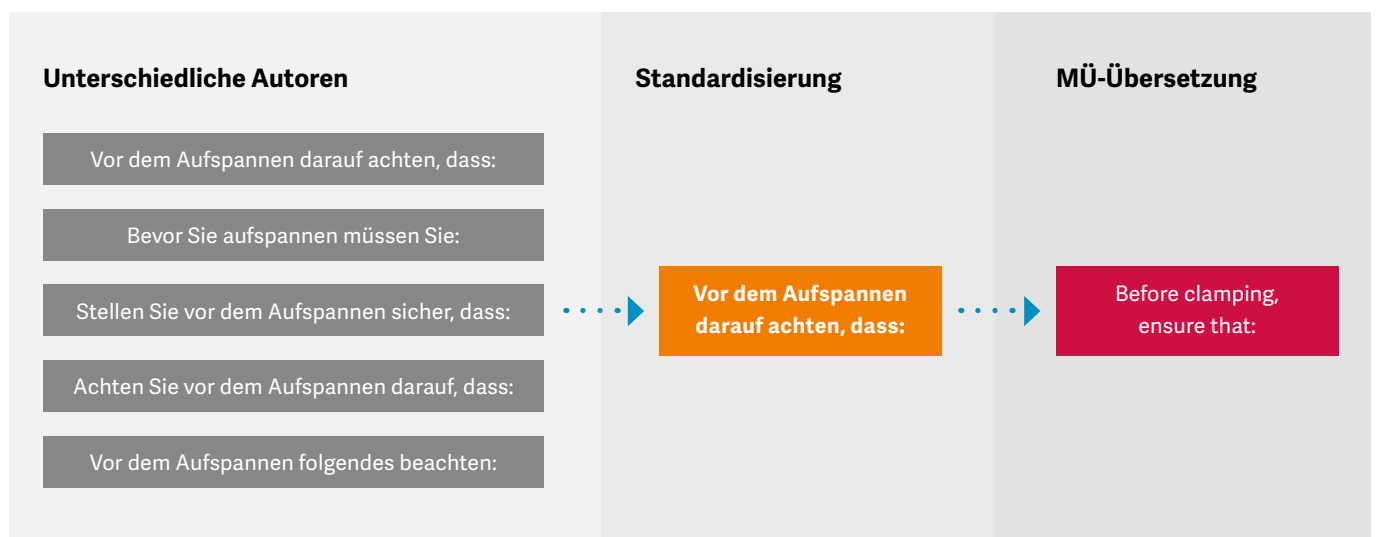
Was können Sie tun, um die Ergebnisse zu optimieren?

Maschinelle Übersetzungssysteme haben Schwierigkeiten mit Texten, die lange Sätze haben, Schreibfehler enthalten, sehr uneinheitlich geschrieben oder ungenau formuliert sind.

Unternehmen können die Ergebnisse von maschineller Übersetzung durchaus beeinflussen, indem sie „besser“ und „maschinengerecht“ schreiben. Das gilt insbesondere bei Firmen, die sehr viel Informationsmaterial produzieren und in denen im Laufe der Jahre mehrere Mitarbeiter Texte verfassen. Dann lohnt es sich, in kontrollierter Sprache zu schreiben. Ein Style Guide fasst die Regeln für optimiertes Schreiben zusammen (Satzlänge, Syntax, Gebrauch der Verben usw.). Sie können die Terminologie Ihres Unternehmens festlegen und definieren, welche Synonyme bevorzugt und welche verboten sind (**Tempomat** und nicht **Geschwindigkeitsbegrenzer**). Dafür eignet sich z. B. das Terminologieverwaltungssystem LookUp der D.O.G. GmbH sehr gut (www.dog-gmbh.de/produkte/lookup).

Hier finden Sie einige Regeln für die Optimierung Ihrer Texte:

1. Prüfen Sie die Rechtschreibung: MÜ-Systeme können mit Schreibfehlern schlecht umgehen.
2. Achten Sie auf korrekte Zeichensetzung im Satz: Kommas, Semikolons, Gedankenstriche, einheitliche Anführungszeichen usw. Sie helfen dem MÜ-System, den Satz richtig zu verstehen.
3. Schreiben Sie einfach: kurze Sätze, wenig Nebensätze.
4. Verwenden Sie eine einheitliche Terminologie, möglichst ohne Oberbegriffe (**Drosselventil** statt **Ventil**): MÜ-Systeme brauchen präzise Benennungen.
5. Verzichten Sie auf Füllwörter (**nämlich, allerdings...**)
6. Verwenden Sie möglichst keine Bezüge auf Wörter außerhalb des Satzes (statt **es, das Telefon** usw.): MÜ-Systeme sind ohne Kontext Informationen verloren.
7. Bei der Formatierung Ihrer Dokumente, achten Sie darauf, dass sie die Arbeit des MÜ-Systems nicht stört: z. B. Hard>Returns mitten im Satz.





Terminologie für Maschinen

Eine gut gepflegte Firmenterminologie ist beim Einsatz maschineller Übersetzungssysteme ein entscheidender Vorteil. Während sich bisher die Terminologearbeit auf die Kommunikation zwischen oder mit Menschen konzentrierte, eröffnet MÜ eine neue Dimension: Terminologie für Maschinen. Dadurch dass MÜ-Systeme zum Teil ganz andere Fehler machen als Menschen, muss die

Terminologearbeit dies berücksichtigen. Mithilfe der erfassten Terminologie können Qualitätssicherungssysteme wie ErrorSpy falsche Termini erkennen, die unter Umständen ein Mensch niemals verwenden würde. So können diese typisch von Maschinen falsch übersetzten Termini in die Terminologie aufgenommen und mit einem Verwendungsattribut wie „verboten“ versehen werden. Darunter fallen u.U. Wörter der allgemeinen Sprache sowie auch Eigennamen oder Produktnamen.

▼ **training** Begriff 35

Definition: **Training a model simply means learning (determining) good values for all the weights and the bias from labeled examples. In supervised learning, a machine learning algorithm builds a model by examining many examples and attempting to find a model that minimizes loss; this process is called empirical risk minimization.**

Bild:

Link: <https://developers.google.com/machine-learning/crash-course/descending-into-ml/training-and-loss>

▼ English

▼ **training**
 Satus: **Geprüft**
 Kontext: **NMT requires substantial amounts of bilingual data for effective model training.**

▼ German

▼ **Ausbildung**
 Satus: **Geprüft**
 Verwendung: **Verboten**
 Benennungskommentar: **Falsch im Kontext von maschinellem Lernen**
 MÜ-Fehler:

▼ **Training**
 Satus: **Geprüft**
 Verwendung: **Erlaubt**

Abbildung 6: Terminologieeintrag zur Prüfung von MÜ-Fehlern



Besseres Trainingsmaterial

Aber nicht nur das optimierte Schreiben von Texten verbessert die MÜ-Ergebnisse, sondern auch die Qualität des Materials, das für das Trainieren eines Modells verwendet wurde. Nach dem Motto „Garbage in, garbage out“ sind die Systeme nur so gut, wie das verwendete Trainingsmaterial.

Viele Unternehmen, die regelmäßig übersetzen lassen, haben möglicherweise Zugriff auf Translation-Memorys (TM). TMs sind Datenbanken mit bereits übersetzten Sätzen, die ihre Übersetzer normalerweise erstellen. Diese Translation-Memorys bilden nach Optimierung der Inhalte eine hervorragende Quelle für das firmenspezifische Trainieren von MÜ-Systemen.

Kosten- und Zeitvorteile

Die am häufigsten genannten Gründe für maschinelles Übersetzen sind Zeit- und Kostenersparnis. Tatsächlich ist für einige Leute allein der Zeitfaktor Grund genug, um maschinell übersetzen zu lassen. Wenn ein Webshop 4 Wochen früher in englischer, französischer oder spani-

scher Sprache verfügbar ist, kann der Verkauf 4 Wochen früher starten. Und das kann viel Wert sein.

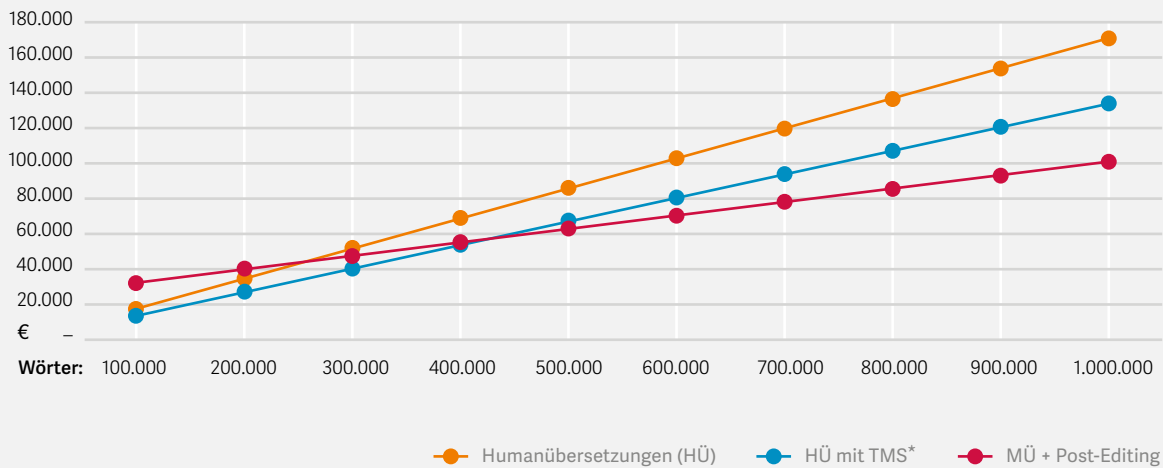
Im Internet kursieren alle möglichen Kostenangaben, so dass es für den Laien nicht immer einfach ist, seriöse von zweifelhaften Aussagen zu trennen. Kosten entstehen hauptsächlich:

- ✓ für den Aufbau eines Modells und anschließend für die regelmäßige Wartung des trainierten Modells einschließlich der technischen Infrastruktur
- ✓ für die Nachbearbeitung (das Post-Editing) von maschinellen Übersetzungen.

Viele Anbieter berechnen einen niedrigen Grundpreis pro übersetztes Wort für die erste Kostenkomponente und einen Wortpreis für das Post-Editing, abhängig vom Umfang der erforderlichen Überarbeitung. Die Kosten können um 20 – 50 % niedriger als bei herkömmlichen Übersetzungen. Wie groß die Einsparung am Ende ist, hängt davon ab, wie anspruchsvoll die Texte sind und ob durch den Einsatz von Translation-Memorys bereits einiges gespart wurde.

Beispiel: Kostenvergleich

Eine Sprache bei 30 % Wiederholungen



* TMS = Translation-Memory-System



Wofür ist MÜ geeignet?

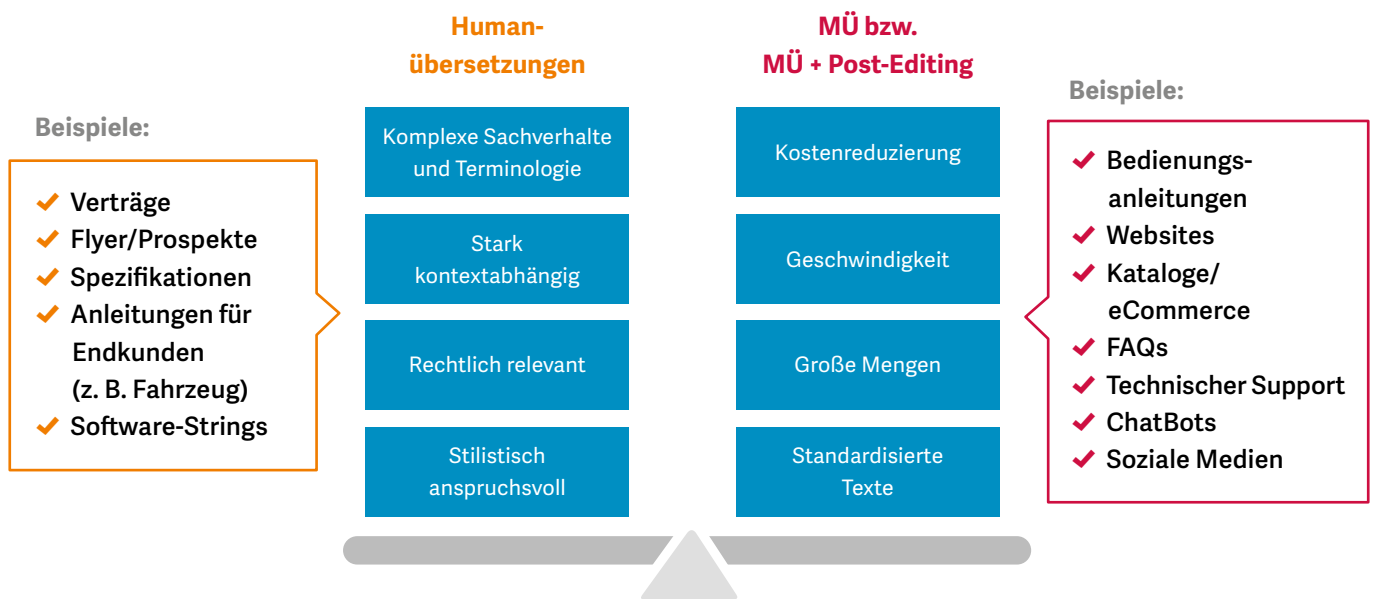
Die Entscheidung über MÜ ist keine Alles-oder-Nichts-Entscheidung. Firmen, die sich für MÜ entscheiden, lassen nach wie vor einen Teil ihrer Übersetzungen klassisch durch Humanübersetzer erstellen. Welche Dokumente oder Informationen für MÜ geeignet sind, hängt von mehreren Faktoren und Prioritäten ab. In der folgenden Grafik haben wir einen Teil dieser Entscheidungsfaktoren zusammengefasst.

MÜ erfolgt für viele Texte und Publikationen meistens in Kombination mit Post-Editing. Überall dort, wo die Texte umfangreich und die Sätze in einem einheitlichen Stil geschrieben sind, sind die MÜ-Ergebnisse, insbesondere von

mit eigenen Texten trainierten MÜ-Systemen relativ gut. Bedienungsanleitungen, Kataloge und Webshops, Websites und Schulungsunterlagen sind einige Beispiele dafür.

Die beste Lösung ist, sich auf einen Übersetzungsdienstleister zu verlassen, der sowohl klassische als auch maschinelle Übersetzungen mit einem trainierten Übersetzungssystem anbietet. Die D.O.G. GmbH hat sich darauf spezialisiert.

Wenn Sie mehr erfahren möchten, können Sie unsere [Website besuchen](#) und lesen, wie wir für unsere Kunden MÜ-Systeme trainieren und einsetzen. Natürlich können Sie sich auch direkt an uns wenden und ein Angebot anfordern oder ein Beratungsgespräch vereinbaren.





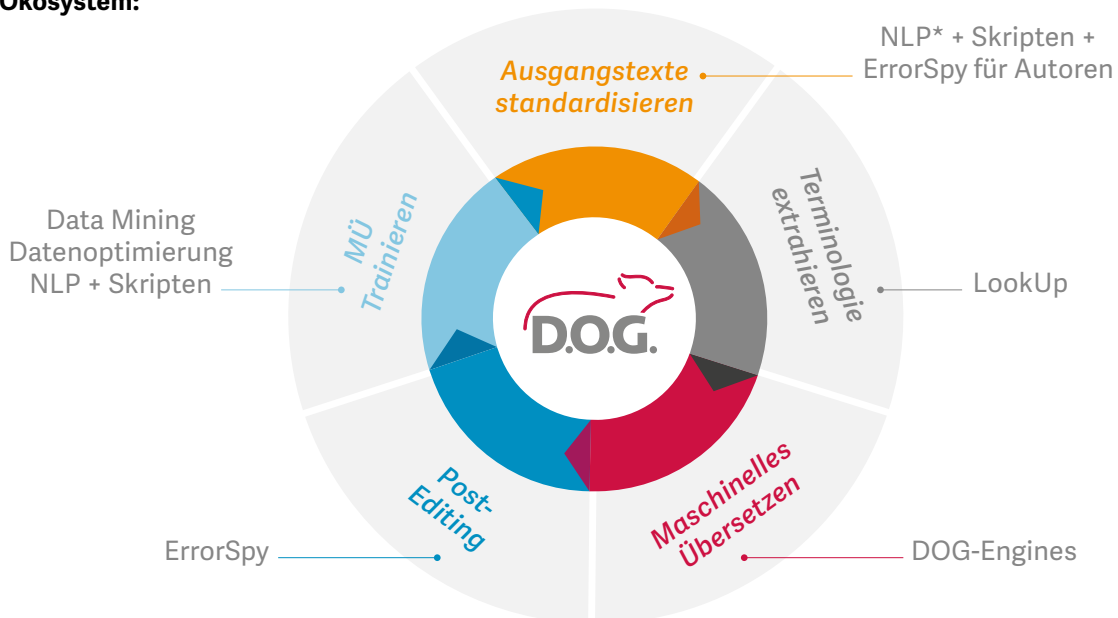
Das D.O.G.-Angebot in Kürze

- ✓ Wir beraten Sie über die Vor- und Nachteile verschiedener Alternativen.
- ✓ Wir arbeiten mit Ihnen ein Pflichtenheft aus, in dem Systeme, Workflows, Integrationsbedarf, Qualitätsrichtlinien festgelegt sind.
- ✓ Wir setzen das für Sie geeignete MÜ-System ein, z. B. ein Neuronales Maschinenübersetzungssystem (NMT), das wir mit Ihren Daten trainieren.
- ✓ Wir bauen das Qualitätsmanagement für Ihre MÜ-Texte auf und setzen ein Team von Post-Editoren ein. Unsere Post-Editoren korrigieren die

maschinell erzeugten Texte nach vereinbarten Kriterien. Dabei verwenden wir unsere Qualitätssicherungssoftware und eine mit Ihnen vereinbarten Terminologie, die wir auf Wunsch gemeinsam mit Ihnen in unserem Terminologieverwaltungssystem LookUp pflegen.

- ✓ Wir pflegen fortlaufend die Sprachressourcen wie Translation-Memories und Terminologie, die für ein optimales Training des Übersetzungssystems wichtig sind.
- ✓ Sie haben einen festen D.O.G. Ansprechpartner, der ein Team von Entwicklern, Post-Editoren und Übersetzern für Ihre Projekte koordiniert.

D.O.G. Leistungen und Ökosystem:



* NLP = Natural Language Processing



Weiterführende Literatur:

- ✓ DIN ISO 18587:2018-02 Übersetzungsdienstleistungen – Posteditieren Maschinell Erstellter Übersetzungen – Anforderungen (ISO 18587:2017).
Beuth Verlag GmbH, Berlin.
- ✓ Koehn, Philipp (2020). Neural Machine Translation.
Cambridge University Press.
doi.org/10.1017/9781108608480
- ✓ Massion, François. 2020. DeepL und Terminologie.
Edition - Deutscher Terminologie-Tag e.V. (DTT) – S. 18–25.
- ✓ Porsiel, Jörg (Hrsg). 2020. Maschinelle Übersetzung für Übersetzungsprofis.
BDÜ Weiterbildungs- und Fachverlags GmbH, Berlin.

Dr. François Massion

01.10.2020

Ihr Ansprechpartner:



Dr. François Massion

Geschäftsführer

E-Mail: francois.massion@dog-gmbh.de

Tel. +49 (0)7152 35411-0



Dokumentation ohne Grenzen GmbH
Neue Ramtelstr. 12
D-71229 Leonberg
Tel. +49 (0)7152 35411-0
Fax +49 (0)7152 35411-50
www.dog-gmbh.de

E-Mail: info@dog-gmbh.de
Tel. +49 (0)7152 35411-11 bzw. -0

