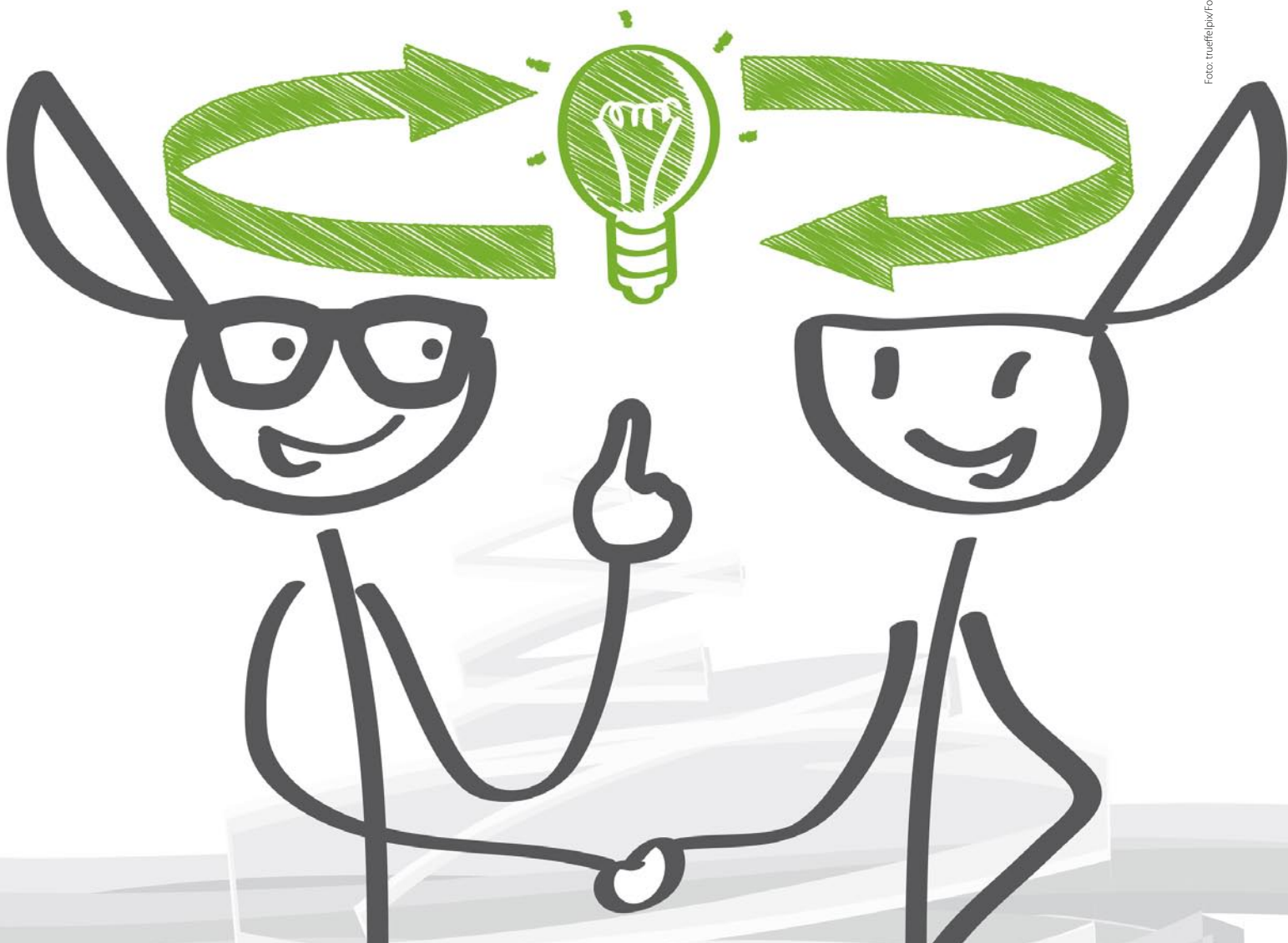




Wie finden Übersetzer und Technische Redakteure zusammen?

Annäherungsversuch

Wenn eine Übersetzung bei Auftraggebern und Übersetzern gleichermaßen Frust hervorruft, liegt es häufig auch daran, dass die Erfordernisse eines modernen Übersetzungsprozesses bei der Erstellung des Ausgangstexts nicht berücksichtigt wurden. Ein Beitrag von Dr. Nicole Keller, mit dem Übersetzer ihren Auftraggebern (nicht nur in der technischen Redaktion) wertvolle Tipps für eine zeit-, kosten- und nervensparende Vorbereitung an die Hand geben können.

Immer wieder ärgern sich Übersetzer über (aus ihrer Sicht) schlecht geschriebene Ausgangstexte und deren unzulänglichen Formatierungen. Diese Kritik ist sicherlich oft gerechtfertigt, aber in den meisten Fällen ist den Technischen Redakteuren kein Vorwurf zu machen, da ihnen bei der Texterstellung gar nicht bewusst ist, dass eine spätere Übersetzung geplant ist und deshalb bestimmte Grundvoraussetzungen erfüllt sein sollten. Darüber hinaus kennen verständlicherweise die wenigsten Redakteure die Kernfunktionalitäten und Funktionsweisen eines Translation-Memory-Systems (TM-Systems), das inzwischen zum Alltag eines Übersetzers, aber nicht eines Technischen Redakteurs gehört. Dieser Artikel soll etwas Licht ins Dunkel bringen und Übersetzer sowie Technische Redakteure näher zusammenrücken lassen.

Die Lösung für die gerade beschriebene Problematik heißt „übersetzungsgerechtes Schreiben“. Wikipedia gibt hierzu eine anschauliche Definition:

Übersetzungsgerechtes Schreiben bezeichnet die Berücksichtigung einer späteren Übersetzung schon bei der Erstellung des Ausgangsdokuments, um mögliche Übersetzungsprobleme zu vermeiden.

Zum übersetzungsgerechten Schreiben gehören neben Verständlichkeit und Konsistenz des Ausgangstextes insbesondere die Kenntnisse über die eingesetzten maschinellen Übersetzungswerkzeuge. Je besser ein Text auf eine automatisierte Verarbeitung zugeschnitten ist, desto schneller, preiswerter, besser und einheitlicher wird eine Übersetzung in eine beliebige Sprache erfolgen können.

Verständlichkeit und Konsistenz sollen an dieser Stelle nicht zur Sprache kommen; der Fokus dieses Beitrags liegt vielmehr auf den Punkten, die formaler Natur sind und deshalb oft zu Problemen führen, wenn Übersetzungen mithilfe eines Translation-Memory-Systems erstellt werden. Screenshots aus verschiedenen TM-Systemen sollen die Problematik vor allem denjenigen veranschaulichen, die normalerweise nicht mit einem solchen System im Arbeitsalltag konfrontiert werden. Das jeweilige System wurde zufällig zur Darstellung ausgewählt, die Auswahl stellt keine Wertung dar.

Absatzmarke vs. Zeilenumbruch

Das erste Problem kann verschiedenste Ursachen haben und ist darauf zurückzuführen, dass Texte bei der Erstellung oder Konvertierung aus PDF-Dateien nicht sorgfältig für eine Bearbeitung in einem Translation-Memory-System vorbereitet werden.

Ein grundlegendes Prinzip vom TM-Systemen ist die Segmentierung des Quelltextes beim Import. Autoren sollten wissen, dass TM-Systeme einen Text zunächst in

Segmente unterteilen. Ein Segment entspricht in der Regel einem Satz. Zum Erkennen dienen die sogenannten Segmentendebegrenzer, die normalerweise Satzzeichen wie einem Punkt, Ausrufezeichen oder Fragezeichen, aber auch einer Absatzmarke (¶) entsprechen. Letztere wird benötigt, um beispielsweise Überschriften, Aufzählungspunkte etc. als eigene Einheit erkennen zu können, wenn am Ende überhaupt kein Satzzeichen steht.

Werden also Sätze grundlos mit einer Absatzmarke (Enter-Taste) in der Mitte getrennt, so wird der Text innerhalb des TM-Systems in unterschiedlichen Segmenten angezeigt (siehe Abb. 1, links). Das Hauptproblem liegt in diesem Fall vor allem darin, dass sehr viele Systeme an solchen Stellen keine manuelle Korrektur der Segmentgrenzen mehr zulassen und der Übersetzer oft nur noch mehr oder weniger sinnvolle Teilsätze übersetzen und im TM speichern kann. Solche Übersetzungseinheiten sind praktisch wertlos, da es sehr unwahrscheinlich ist, dass sie so oder so ähnlich (also mit derselben Trennung) noch einmal in einem Text vorkommen. Ist derselbe Satz bzw. dasselbe Segment noch einmal im Text enthalten, wird keine hundertprozentige Übereinstimmung angezeigt und der Übersetzer übersetzt den Satz noch einmal. Das kostet Zeit und damit auch Geld. Und das Ergebnis ist womöglich nicht mit der ersten Übersetzung konsistent.

Dieses Phänomen kommt sehr oft vor, wenn eine PDF-Datei zum Übersetzen geliefert wird. Dieser Text muss zunächst in eine Word-Datei konvertiert werden, da TM-Systeme grundsätzlich nur Word-Dateien sauber einlesen. Bei der Umwandlung wird am Zeilenende sehr häufig eine Absatzmarke gesetzt, ganz gleich, ob hier ein Satz zu Ende ist oder nicht. Um eine falsche Segmentierung zu vermeiden, sollten diese falsch gesetzten Absatzmarken grundsätzlich zunächst aus dem Dokument entfernt werden, bevor es eingelesen wird.

Derselbe Effekt tritt auf, wenn Redakteure Sätze aus optischen Gründen in der Mitte trennen (z.B. bei Power-Point-Präsentationen). Solche Umbrüche innerhalb eines Satzes sollten grundsätzlich mit einem Zeilenumbruch (↵) durchgeführt werden, da dieser von einem TM-System nicht als Segmentende interpretiert wird. Da dies bei der Quelltexterstellung zunächst unerheblich ist, berücksichtigen viele Autoren diese Vorgehensweise nicht. Als Übersetzer sieht man sich dann jedoch mit unnötigen Problemen konfrontiert, die während der Übersetzung oft nicht mehr zu lösen sind.

Aber auch der entgegengesetzte Fall führt zu Problemen bei der Übersetzung. Wird ein Zeilenumbruch statt einer Absatzmarke verwendet, um beispielsweise für einen zweiten Aufzählungspunkt in die nächste Zeile zu wechseln, werden beide Sätze als ein Segment und damit als eine Ein-

heit angezeigt und bilden daher keine gute Grundlage für eine sinnvolle Übersetzungseinheit (siehe Abb. 1, rechts). In diesen Fällen können die Segmente jedoch während der Übersetzung manuell getrennt werden.

Absatzmarke	Zeilenumbruch
Ein Satz sollte nie mit einem harten ¶	--Der erste Aufzählungspunkt--
Return in der Mitte getrennt werden. ¶	--Der zweite Aufzählungspunkt--
1 Ein Satz sollte nie mit einem harten 2 Return in der Mitte getrennt werden.	4 --Der erste Aufzählungspunkt-- --Der zweite Aufzählungspunkt--

Abb. 1: Beispiel einer falsch gesetzten Absatzmarke (links) und eines falsch gesetzten Zeilenumbruchs (rechts) in MS Word und das entsprechende Ergebnis in einem TM-System (SDL Trados Studio)

Nebenbei kann die Konvertierung von PDF-Dateien noch weitere Fehler mit sich bringen. Oft gehen Formatierungen verloren oder es wird sogar die Reihenfolge des Textes verändert, wenn der Originaltext zum Beispiel in einem Desktop-Publishing-Programm erstellt wurde. Schon aus diesem Grund sollten Texte zur Übersetzung immer im Originalformat vorliegen, da dies in einem TM-System einfacher zu bearbeiten und weniger vorbereitungsintensiv ist.

Sinnloserweise unterstützen inzwischen viele Translation-Memory-Systeme den direkten Import von PDF-Dateien, weil diese Funktion immer wieder von Anwendern gewünscht wurde. Von der Nutzung ist allerdings in den meisten Fällen abzuraten, da das Ergebnis in der Regel unbrauchbar ist und nicht sinnvoll mit einem TM-System übersetzt werden kann.

Leerzeichen vs. Tabulator

Das Einrücken von Texten oder Textpassagen wird leider immer noch sehr häufig „manuell“ mit Fix-Tabulatoren oder sogar Leerzeichen vorgenommen, obwohl es in jeder Anwendung dafür bestimmte Funktionen gibt. Solange ein Text nicht übersetzt werden muss, hat diese Vorgehensweise zunächst keinerlei negative Auswirkungen. Sobald ein solcher Text aber in ein Translation-Memory-System importiert wird, sind die Auswirkungen eher unschön und führen auch nach der Übersetzung oft zu schlecht formatierten Texten. Darüber hinaus zeigt Abbildung 2, dass eine solche Vorgehensweise auch zu nicht sehr sinnvollen Segmenten führen kann (linkes Beispiel).



Abb. 2: Auswirkungen von Leerzeichen und Tabulatoren (memoQ)

Übersetzer können diese „Absatzformatierungen“ zwar im TM-System setzen. Da aber nicht in der Originalanwendung gearbeitet wird, lässt sich so nicht einschätzen, inwieweit die meist längere oder kürzere Übersetzung im Zieltext noch in den Textfluss passt. In der Praxis bedeutet dies, dass sich Übersetzer diese Stellen notieren müssen, um sie nach der Übersetzung in der Originaldatei wieder einzuarbeiten. Wird die finale Formatierung nicht vom Übersetzer selbst, sondern beim Auftraggeber vorgenommen, müssen solche Informationen per E-Mail mitgeschickt werden, damit sie am Ende wieder in den Text eingearbeitet werden können.

Zahlen und Zahlenformate

Zahlen und deren verschiedenen Formate und Schreibweisen stellen oft eine große Herausforderung für TM-Systeme dar. Zwar erkennen in der Regel alle Übersetzungssysteme Zahlen oder Zahlenformate (z. B. Datumsangaben oder Kommazahlen) automatisch und können sie auch „übersetzen“ oder zumindest im Rahmen der Qualitätssicherung prüfen. Dies funktioniert aber nur reibungslos, wenn sich die Autoren an die entsprechenden Konventionen der Ausgangssprache halten. In Deutschland wird eine 1.000er-Zahl beispielsweise mit einem Punkt unterteilt, in England dagegen mit einem Komma (1,000). Werden diese Zeichen nun nicht konsistent verwendet oder stattdessen Leerzeichen gesetzt, wird die Zahl vom System nicht oder falsch erkannt und gegebenenfalls auch durch automatische Konvertierungsprozesse falsch „übersetzt“.

Abbildung 3 zeigt, dass Zahlen im falschen Format gar nicht erst als zusammengehörige Einheit erkannt werden (Beispiel: 12.65 Euro), da dieses Format nicht der Vorgabe im Deutschen entspricht. In diesem Fall geht das System davon aus, dass es sich um zwei unabhängige Zahlen, 12 und 65, handelt. Zumindest erhalten die Übersetzer hierauf einen Hinweis: Das System markiert erkannte Einheiten, wie im Beispiel mit blauen Linien. Während die erste Zahl falsch als zwei separate dargestellt wird, ist die Datumsangabe (30.03.2015) korrekt angegeben und wird deshalb auch als eine zusammengehörige Einheit erkannt.

Die Rechnung beläuft sich auf 12.65 Euro. Sie soll bis zum 30.03.2015 bezahlt werden.

Abb. 3: Falsch (12.65 Euro) und korrekt (30.03.2015) geschriebenes Zahlenformat (Across)

Möglich ist jedoch auch, dass Zahlen, die ein falsches Format haben, trotzdem als eine Zahleneinheit erkannt werden. Hier liegt das Risiko darin, dass das System einen falschen Übersetzungsvorschlag liefert.

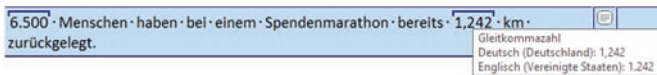


Abb. 4: Korrekt (6.500) und falsch (1,242 km) geschriebenes Zahlenformat mit falschem Übersetzungsvorschlag im Tooltip (1.242)

Terminologie

Neben der konsistenten Verwendung von Termini ist auch deren Schreibweise essentiell. Dies ist einerseits wichtig, damit Termini aus der Terminologiedatenbank richtig bzw. überhaupt erkannt werden. Wird ein Mehrwortterminus beispielsweise einmal mit und einmal ohne Bindestriche geschrieben (Abb. 5), wird die Benennung vom System nicht als identisch erkannt und vom Übersetzer womöglich aus Versehen anders übersetzt. Dies kann auch passieren, wenn nicht mit der Trennfunktion einer Software gearbeitet wird, sondern aus Unwissenheit oder Bequemlichkeit ein Bindestrich zur Trennung von Wörtern verwendet wird.

16	Im weiteren Sinn umfasst der Begriff alle Krankheiten des Herz-Kreislauf-Systems.
19	Im weiteren Sinn umfasst der Begriff alle Krankheiten des Herz Kreislauf Systems.

Abb. 5: Identischer Satz mit unterschiedlicher Schreibweise von Herz-Kreislauf-System. In diesem Fall wird nur die Schreibweise mit Bindestrich erkannt (gelbe Markierung) (Memsources)

Andererseits ist es in der Praxis auch keine Seltenheit, dass im Quelltext Synonyme oder alternative Schreibweisen verwendet werden. Sind diese Alternativen nicht in der Terminologiedatenbank hinterlegt, werden sie auch nicht automatisch angezeigt. Darüber hinaus wird ein 100%-Match (identischer Satz) z. B. nur noch als 85%-Match (siehe Abbildung 6) angezeigt, obwohl die Sätze inhaltlich zu 100 Prozent übereinstimmen.

17	Im engeren Sinn versteht man darunter jedoch vor allem die Erkrankungen, deren gemeinsame pathologische Ursache die Atherosklerose ist.
20	Im engeren Sinn versteht man darunter jedoch vor allem die Erkrankungen, deren gemeinsame pathologische Ursache die Arteriosclerosis ist.
1	Im engeren Sinn versteht man darunter jedoch vor allem die Erkrankungen, deren gemeinsame pathologische Ursache die Atherosklerose ist.
85	In the narrower sense it refers to particular diseases, commonly caused by atherosclerosis.

Abb. 6: Identischer Satz mit alternativer Schreibweise und 85%-Match aus dem Translation-Memory (Memsources)

Fazit

Die gezeigten Beispiele belegen, dass es neben inhaltlichen Herausforderungen zahlreiche formale Aspekte gibt, die den Übersetzungsprozess in einem Translation-Memory-System verzögern können. Je nach System können diese Fehler während der Übersetzung nicht behoben werden, weil zum Beispiel Segmentgrenzen nicht angepasst werden können oder der Quelltext nicht verändert werden kann. Dies führt zwangsläufig zu schlechten oder sinnlosen Übersetzungseinheiten im Translation-Memory und zu einer schlechteren Analyse nachfolgender Texte.

Sind Segmente nicht sauber im TM gespeichert, da sie auseinandergerissen sind oder durch überflüssige Tags (z. B. Tabulatoren oder Leerzeichen) „verunreinigt“ wurden, ergibt die Analyse eines neu zu übersetzenden Textes bei identischen Segmenten meist einen Wert unterhalb der 100%-Grenze.

Auch die inkonsistente Verwendung von Termini im Quelltext führt zu weniger 100%-Matches und einer vermeidbaren inkonsistenten Verwendung von verschiedenen Übersetzungen im Zieltext.

Unterm Strich führen solche Fehler immer dazu, dass die Übersetzungskosten und die für die Übersetzung benötigte Zeit unnötig in die Höhe getrieben werden – Zeit, die sinnvollerweise in die Übersetzung selbst investiert werden sollte. Fakt ist also: Je sauberer ein Text formatiert wurde, umso mehr können sich die Übersetzer auf das eigentliche Übersetzen konzentrieren, was automatisch auch die Qualität der Übersetzung steigert. ■

Dr. Nicole Keller

Nicole Keller ist Dozentin am Institut für Übersetzen und Dolmetschen der Universität Heidelberg. Schwerpunkte ihrer Arbeit sind Translation-Memory-Systeme und Terminologiedatenbanken. Sie arbeitet als Trainerin und Beraterin für computergestützte Übersetzungswerkzeuge sowie als freiberufliche Übersetzerin und betreut seit Januar 2013 als Rubrikleiterin das Ressort Werkzeuge im MDÜ. Kontakt: mail@nicole-keller.com

