

Die 10 goldenen Regeln für die Erstellung übersetzungsgerechter Quelldokumente

Der vorliegende Leitfaden soll Ihnen als Orientierung dienen, um Quelldokumente bereits bei der Erstellung optimal für die nachfolgende Übersetzung vorzubereiten. Ein gut vorbereitetes und sauber formatiertes Quelldokument kann bei der Übersetzung mit einem Translation Memory-System (TM-System) viel Zeit und Kosten einsparen, da die Wiedererkennungsfunktionalitäten des TM-Systems nur dann sinnvoll greifen, wenn die zu übersetzenden Einheiten auch tatsächlich identisch oder ähnlich sind.

Die folgenden Punkte sind aus der praktischen Arbeit mit Across hervorgegangen. Sie mögen für den Autor eines Textes auf den ersten Blick eher unerheblich erscheinen, sie stellen für den Übersetzer aber teilweise große Probleme dar.

1 PDF-Dateien vs. Originaldateiformate

Es sollte, wenn möglich, vermieden werden, PDF-Dateien als Quelldokumentformat für die Übersetzung zur Verfügung zu stellen. Versuchen Sie stets die Originaldateiformate, die als Grundlage für die Erstellung der PDF-Dateien dienten, bereit zu stellen, da PDF-Dateien derzeit nicht unmittelbar verarbeitet werden können, sondern vor der Übersetzung in ein anderes Format (z.B. Word) umgewandelt werden müssen. Die umgewandelten Dokumente müssen in der Regel vor der Übersetzung noch einmal bearbeitet werden, da der konvertierte Text meist zu viele Formatierungsmängel aufweist, um ihn sinnvoll mit einem TM-System übersetzt werden zu können. Diese Bearbeitung ist immer mit einem zusätzlichen Zeit- und somit auch Kostenaufwand verbunden und verzögert den Beginn der Übersetzung.



2 Harte Zeilenumbrüche

Vermeiden Sie harte Zeilenumbrüche (Absatzmarken) innerhalb von Sätzen, da ansonsten keine sinnvollen Segmente für die Übersetzung angeboten werden können. Zeilenumbrüche sollten nur verwendet werden, wenn tatsächlich ein neuer Absatz beginnen soll. TM-Systeme entscheiden anhand von Segmentendebegrenzern, wo eine Übersetzungseinheit (normalerweise ein Satz) zu Ende ist. Diese Zeichen sind in der Regel ., !, ? und ¶.

Ein Zeilenumbruch wird also stets als Segmentende erkannt und es bedarf einer manuellen Nachbearbeitung, wenn der Zeilenumbruch innerhalb eines Satzes steht und diesen folglich in zwei Segmente unterteilt. Die manuelle Anpassung durch den Übersetzer ist ein zusätzlicher Zeitaufwand, gleichzeitig ergibt eine vorgeschaltete Analyse auch weniger Suchtreffer im TM (Matches), wodurch eine Übersetzung gegebenenfalls unnötig teurer wird.

Sehr häufig werden harte Zeilenumbrüche in PowerPoint oder in DTP-Programmen falsch gesetzt, weil die Kenntnisse in der Arbeit mit diesen Systemen nicht ausreichend sind.

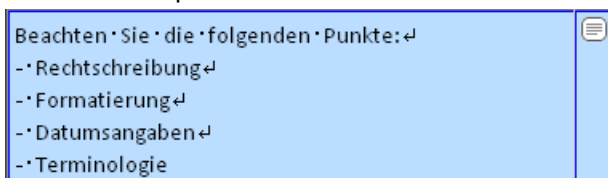
Beispiel für die Darstellung eines Satzes in crossDesk, der einen falschen Zeilenumbruch enthält:

Dieses • erste • Beispiel • hat • mitten • im • Satz • eine • Absatzmarke, • da • der • Text • entweder • aus • einer • PDF-Datei •	
herauskopiert oder aus Formatierungsgründen einfach eine Absatzmarke gesetzt wurde. Der Satz wird dadurch an einer unlogischen Stelle auseinandergerissen und muss vom Übersetzer manuell nachbearbeitet werden.	

3 Weiche Zeilenumbrüche

Weiche oder manuelle Zeilenumbrüche (**Strg+Enter**) sollten ebenfalls vermieden werden. Sie werden von TM-Systemen nicht als Segmentende interpretiert, weshalb solche Einheiten nicht korrekt erkannt werden und vom Übersetzer manuell nachbearbeitet werden müssen.

Das folgende Textbeispiel enthält am Ende der Aufzählungspunkte jeweils einen weichen Zeilenumbruch, wodurch der komplette Textbereich als eine Einheit zur Übersetzung im crossDesk angeboten wird.



Weiche Zeilenumbrüche werden häufig unbeabsichtigt durch das Kopieren von Texten zwischen verschiedenen Anwendungen in Quelldokumente eingefügt. Dies kommt zum Beispiel sehr häufig vor, wenn der zu übersetzende Text z.B. aus einer E-Mail in ein Word-Dokument kopiert wird.

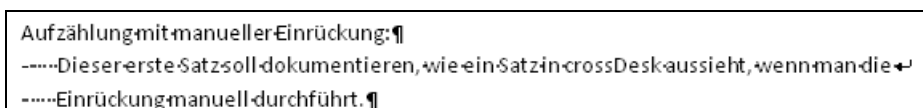
4 Manueller Seitenumbruch

Sehr häufig werden in Texten Seitenumbrüche eingefügt, weil zum Beispiel eine Überschrift nicht am unteren Seitenrand, sondern auf der nächsten Seite oben erscheinen soll. Für die Verbesserung des Layouts und eine bessere Lesbarkeit des Textes fügen Autoren deshalb oft manuelle Seitenumbrüche ein. Während der Übersetzung wird ein Text normalerweise länger oder kürzer als der Ausgangstext, sodass es eher unwahrscheinlich ist, dass solche Seitenumbrüche im Zieltext an derselben Stelle stehen. Normalerweise werden diese manuellen Seitenumbrüche während der Übersetzung übersprungen und damit nicht im internen Übersetzungseditor in den Zieltext eingefügt, sondern am Ende, wenn das Dokument wieder in das ursprüngliche Dateiformat konvertiert wurde.

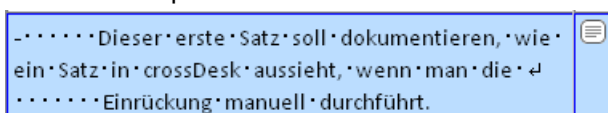
5 Leerzeichen und Tabulatoren zur Formatierung

Versuchen Sie das Einrücken von Texten mit Tabulatoren oder Einzügen vorzunehmen und verwenden Sie hierfür keinesfalls mehrere Leerzeichen. Nach dem Einlesen des Dokuments in ein TM-System werden diese Zeichen allesamt angezeigt und die Darstellung im Übersetzungseditor ist einerseits verwirrend, andererseits ist eine solche manuelle Anpassung auf die Quellsprache ausgerichtet. Die Übersetzung mit denselben Leerzeichen sieht in 99% der Fälle anders aus als im Quelldokument. Der Text muss also nach der Übersetzung in jedem Falle nachbearbeitet werden.

Beispieltext in Word:



Derselbe Beispieltext im crossDesk:



Ein Übersetzer kann bei seiner Arbeit oft nur schwer beurteilen, ob die Leerzeichen tatsächlich eine bestimmte Funktion haben oder nur Formatierungszwecken dienen. Wenn wir z.B. davon ausgehen, dass der Übersetzer die Leerzeichen nicht aus der Übersetzung löscht und versucht, sie im Zieltext an denselben Stellen zu platzieren wie im Quelldokument, dann würde die Übersetzung nach dem Export folgendermaßen aussehen:

Bulleted list with manual indentation:

- This first sentence will demonstrate how a sentence will look like in crossDesk if you manually set spaces instead of defining an indentation.

6 Datums-, Zeit- und Zahlenformate

Bei der Erkennung von Datums-, Zeit- und Zahlenformaten orientieren sich TM-Systeme an vorgegebenen Regeln. So ist in den Systemeinstellungen von Across z.B. hinterlegt, dass Datumsangaben im Deutschen die Formate TT.MM.JJJJ oder TT.MM.JJ haben. Steht nun zwischen den Zahlen ein Leerzeichen, wird die Datumsangabe nicht mehr als zusammenhängende Zahlengruppe erkannt und kann daher nicht in der Übersetzung auf die korrekte Verwendung hin überprüft werden. Sehr häufig geschieht dies z.B. bei Datumsangaben oder Tausenderzahlen.

Beispiel: 08.10.2010 vs. 08. 10. 2010

Datumsangabe · 08. · 10. · 2010 · vs. · 08.10.2010 ·

Beispiel: 5.000 vs. 5 000

· Beispiele: · 5.000 · vs. · 5 · 000 ·

Die blauen Linien weisen den Übersetzer darauf hin, dass eine Zahl vorliegt und welche Zahlenbereiche als zusammenhängende Einheiten erkannt wurden. Ist die blaue Linie unterbrochen, wurden mehrere Einheiten erkannt. In diesem Fall kann keine sinnvolle Überprüfung auf die korrekte Übernahme der Zahlenformate durchgeführt werden.

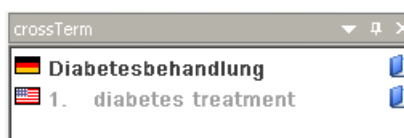
Es ist somit empfehlenswert, dass bei Texten, die viele Datums-, Zeit- und Zahlenformate enthalten, zunächst eine einheitliche Schreibweise festgelegt und diese konsistent verwendet wird.

7 Einheitliche Schreibweise von Fachwörtern

Die unterschiedliche Schreibweise von Fachwörtern, z. B. die Schreibung in einem Wort (Diabetesbehandlung), in zwei Wörtern (Diabetes Behandlung) oder die Schreibung mit Bindestrich (Diabetes-Behandlung), ist häufig die Ursache für eine inkonsistente Übersetzung, da die automatische Terminologieerkennung in diesen Fällen nicht greift.

Beispiel für korrekte Schreibweise:

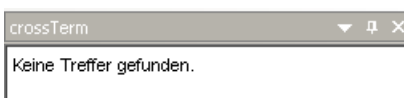
Der erste Satz enthält die korrekte Schreibweise von „Diabetesbehandlung“. Im zweiten Satz werden zwei andere Schreibweisen verwendet: Diabetes Behandlung und Diabetes-Behandlung.



Die rote Markierung zeigt an, dass der Term in der Terminologiedatenbank crossTerm gespeichert vorliegt. Dem Übersetzer wird die hinterlegte Übersetzung vorgeschlagen. Der Übersetzer kann diesen Vorschlag direkt in die Übersetzung übernehmen.

Beispiel für alternative/falsche Schreibweise:

Der erste Satz enthält die korrekte Schreibweise von „Diabetesbehandlung“. Im zweiten Satz werden zwei andere Schreibweisen verwendet: Diabetes Behandlung und Diabetes-Behandlung.



Das System erkennt das Fachwort nicht und das Terminologiefenster bleibt leer. Der Übersetzer weiß also in diesem Fall nicht, dass es bereits verwertbare Terminologieinformationen gibt und übersetzt im Zweifelsfall inkonsistent oder falsch.

8 Korrekte Schreibweise von Abkürzungen

Verwenden Sie immer die korrekte Schreibweise von Abkürzungen. Falls Sie eine unübliche oder falsche Schreibweise einer Abkürzung verwenden, wird die Segmentierung während der Übersetzung nicht mehr richtig sein. Zum Beispiel ist die offizielle Abkürzung für gegebenenfalls ggf. Man findet aber auch immer wieder folgende Abkürzungen: ggb., ggbfs., ggf., ggfs., ggfls, ggfs, ggfls. In Across ist nur ggf. als Abkürzung hinterlegt; bei der Verwendung aller anderen Abkürzungen wird der Satz in zwei Segmente geteilt, die vom Übersetzer zunächst manuell angepasst werden müssen.

9 Überflüssige Formatierungen

Wenn Sie beispielsweise häufig mit farbigen Markierungen im Text arbeiten, um Textstellen optisch hervorzuheben, sollten Sie darauf achten, dass diese vor der Übersetzung vollständig entfernt wurden und nicht noch z. B. Zeilenumbrüche oder Leerzeichen vorhanden sind, denen diese Formatierung noch zugewiesen ist. Diese „unsichtbaren“ Formatierungen werden dem Übersetzer ansonsten als mögliche Formatierungen angeboten und führen dazu, dass ein Übersetzer gegebenenfalls direkt in der falschen Schriftart/-farbe schreibt.

Beispiel aus Word:

Am Satzendewurde das Absatzzeichen nicht in derselben Formatierung wie der Rest des Satzes formatiert ¶

10 Silbentrennung

Wenn Sie in Ihrem Text die Silbentrennung verwenden möchten, achten Sie darauf, dass Sie entweder mit der „automatischen Silbentrennung“ arbeiten oder den bedingten Trennstrich manuell einfügen. Viele Benutzer fügen in den meisten Fällen nur einen einfachen Bindestrich ein, da er optisch dasselbe Ergebnis liefert. In diesem Fall führt das innerhalb des Translation-Memory-Systems zu folgenden Problemen:

Normale Bindestriche werden als zusätzliches Zeichen im Wort erkannt und für die Übersetzung ebenfalls als solches dargestellt. Das bedeutet einerseits, dass die Übersetzungseinheit, die im TM gespeichert wird, keinen 100%-Match ergibt, wenn derselbe Satz noch einmal ohne Bindestrich vorkommt. Andererseits wird die automatische Terminologieerkennung für diese Fälle keinen Treffer aus der Terminologiedatenbank anzeigen, da die Benennung durch ein falsches Zeichen getrennt wurde.

Erster Beispielsatz mit korrekter Silbentrennung in Word:

The first sentence contains an optional hyphen-ation that was inserted manually. ¶

Erster Beispielsatz mit korrekter Silbentrennung in Across. Die Benennung wurde vom System erkannt.

The · first · sentence · contains · an · optional · hyphen-ation · that · was · inserted · manually. ¶

Zweiter Beispielsatz mit falscher Silbentrennung in Word:

The second sentence just contains a hyphen in the middle of the word hyphen-ation. ¶

Zweiter Beispielsatz mit falscher Silbentrennung in Across. Die Benennung wurde vom System nicht erkannt.

The · second · sentence · just · contains · a · hyphen · in · the · middle · of · the · word · hyphen-ation. ¶