



Maschinelle Übersetzung

Postediting – Schreckgespenst oder Perspektive?

Maschinelle Übersetzung soll Zeit und Kosten sparen. Für viele Einsatzgebiete kommt sie ohne Feinschliff von Hand jedoch nicht aus. Harald Elsen erläutert, wie Postediting in Verbindung mit einem CAT-Tool funktioniert, wo die Grenzen liegen – und was man als Übersetzer mitbringen sollte, um sich damit eine weitere Facette eines vielschichtigen Berufsbilds zu erschließen.

Was Postediting (PE) ist, gibt der Buchtitel *Repairing Texts* (Krings, 2001) recht prägnant wieder. Vielleicht ist das Wort reparieren nicht völlig angemessen – aber es stellt zumindest eine angemessene Form der Übertreibung dar, um es auf den Punkt zu bringen:

Ein maschinelles Übersetzungssystem (MT-System) produziert mithilfe von großen statistisch ausgewerteten Datenmengen und/oder unterstützt durch linguistische Informationen und Regeln eine Übersetzung. Diese maschinelle Vor-Übersetzung kann den qualitativen Mindestanforderungen entsprechen und zur „Veredelung“ nachbearbeitet (repariert, posteditiert) werden, oder diese maschinelle Vor-Übersetzung wird aufgrund von zu großen Mängeln verworfen und normal übersetzt.

Im letzten Satz sind die Kernaufgaben des Posteditings deutlich beschrieben:

(1) Feststellen, ob ein MT-Segment¹ den Mindestanforderungen entspricht und in angemessener Zeit repariert werden kann;

(2) Nachbearbeiten des MT-Segments, das als verwertbar eingestuft wurde;

(3) gegebenenfalls Übersetzen des Ausgangssegments, für das kein verwertbares MT-Segment existiert.

Damit sich das Postediting wirtschaftlich auszahlt, müssen die Aufgaben (1) und (2) in einem kürzeren Zeitraum durchgeführt werden, als eine normale Übersetzung dauern würde. Zusätzlich muss die gewonnene Zeiteinsparung den Zeitverlust kompensieren, der durch die Verwendbarkeitsbewertung jedes MT-Segments entsteht, unabhängig davon, ob es verwendet wird oder nicht (siehe Tabelle „Dauer der Verwendbarkeitsbewertung und Produktivität“²).

¹ Wir haben uns hier für die Formulierung MT-Segment entschieden, da MÜ-Systeme in den meisten Fällen, ähnlich den Translation-Memory-Systemen, Texte satzbasiert verarbeiten.

² Beide MT-Ergebnisse wurden mit dem MT-System LucyLT der Firma Lucy Software und Services GmbH erstellt: Version DE (MT1) ohne Anpassung des Systems, Version DE (MT2) mit angepassten Lexikoneinträgen und Fachgebetsangaben.



Dauer der Verwendbarkeitsbewertung und Produktivität

EN	It also restores replicated files in the file system to their designated degree of replication when a previous disk failure or removal of a disk has made some replica data inaccessible.	
DE (MT ₁)	Es gibt auch ihrem bestimmten Grad von Erwidern reproduzierte Dateien im File System zurück, wenn ein bisheriger Datenträgerausfall oder Beseitigung einer Scheibe einige Kopiedaten unerreichbar gemacht hat.	<ul style="list-style-type: none"> • Hohe Bewertungszeit • Geringe Qualität • Geringe Verwendbarkeit • Produktivitätsverlust
DE (MT ₂)	Es stellt auch replizierte Dateien im Dateisystem zu ihrem bestimmten Grad von Replikation wieder her, wenn ein vorheriger Datenträgerausfall oder Entfernen eines Datenträgers einige Replikationsdaten unerreichbar gemacht hat.	<ul style="list-style-type: none"> • Geringe Bewertungszeit • Höhere Qualität • Hohe Verwendbarkeit • Produktivitätsgewinn

Ist der Anteil der MT-Segmente vom Typ MT1 hoch und der Anteil vom Typ MT2 gering, dann kann der Produktivitätsgewinn durch MT2 voraussichtlich den Zeitverlust durch MT1 nicht kompensieren. Im umgekehrten Fall ist jedoch die Kompensation des Zeitverlustes sehr wahrscheinlich.

Somit lässt sich die Mindestanforderung für einen wirtschaftlich sinnvollen Einsatz von Postediting folgendermaßen skizzieren:

$$\begin{matrix} \text{Aufwand für} \\ \text{Verwertbarkeitsbewertung aller Segmente +} \\ \text{Nachbearbeitungszeit verwertbarer} \\ \text{MT-Segmente +} \\ \text{Normale Bearbeitungszeit der übrigen Segmente} \end{matrix} < \begin{matrix} \text{Aufwand für} \\ \text{normale Bear-} \\ \text{beitung aller} \\ \text{Segmente} \end{matrix}$$

Die Wirtschaftlichkeit des Posteditings hängt also hauptsächlich von drei Faktoren ab (wobei die „2 Second Rule“ mittlerweile nicht mehr als hilfreich gilt, sondern eher als Aufforderung verstanden wird, dass für die Verwertbarkeitsbewertung so wenig Zeit wie möglich investiert werden soll):

Faktor	Beschreibung
Dauer der Verwertbarkeitsbewertung	Die Verwertbarkeitsbewertung muss möglichst kurz sein. Aus diesem Grund wird oft als Vorgabe die „2 Second Rule“ genannt, nach der die Verwertbarkeitsbewertung nicht länger als zwei Sekunden dauern soll. Kann innerhalb dieser zwei Sekunden die Bewertung nicht erfolgen, dann soll das MT-Segment verworfen werden.
Nachbearbeitungs-aufwand verwertbarer MT-Segmente	Die MT-Segmente müssen qualitativ so gut sein, dass ihre Nachbearbeitung – ähnlich einem <i>Fuzzy Match</i> – schneller zu einer zufriedenstellenden Übersetzung führt (Produktivitätsvorteil). Der Nachbearbeitungsaufwand hängt wiederum größtenteils von den Qualitätsvorgaben/-erwartungen ab.
Anzahl verwertbarer MT-Segmente	Die Anzahl verwertbarer MT-Segmente muss so hoch sein, dass sich ein signifikanter Produktivitätsvorteil einstellt.

Für Texte, die vollständig maschinell übersetzt werden, wie z. B. Knowledge-Base-Einträge, können Daten zur Ermittlung der Produktivität relativ leicht erhoben werden, wodurch sich leicht Rückschlüsse auf die Wirtschaftlichkeit von Postediting ziehen lassen.

Als weitere wirtschaftliche Stellschraube wird häufig auch diskutiert, ob – je nach Verwendungskontext und den damit verbundenen Qualitätsanforderungen – unbedingt ausgebildete Übersetzer die Aufgabe des Posteditings übernehmen müssen oder ob nicht anders qualifizierte Ressourcen, wie z. B. Sprachwissenschaftler, diese Aufgabe günstiger ausführen können. Dabei gilt es aber unserer Meinung nach zu bedenken, dass Segmente, für die es keine verwertbaren MT-Segmente gibt, normal übersetzt werden müssen und die voraussichtlich geringe Übersetzungserfahrung solcher Ressourcen die Produktivität und Qualität allgemein beeinträchtigen würde.

Erstes Resümee

PE-Fähigkeiten	Ein Posteditor muss in der Lage sein, die Verwertbarkeit und den Nachbearbeitungsaufwand eines MT-Segments für den jeweiligen Verwendungskontext schnell einzuschätzen.
PE-Richtlinien	Für die Entscheidung über die Verwertbarkeit eines MT-Segments müssen eindeutige Qualitätsvorgaben/-erwartungen für den jeweiligen Verwendungskontext vorliegen (Information, interne Kommunikation, technische Dokumentation, Reparaturleitfaden etc.).
MT-System	Ist der Anteil der nicht verwertbaren MT-Segmente zu hoch (z. B. weil die MT-Ausgabe für den jeweiligen Verwendungskontext zu schlecht ist), dann wird durch die zu treffenden Verwertbarkeitsentscheidungen so viel Zeit verloren, dass der durch die verwertbaren MT-Segmente erzielte Produktivitätsgewinn verloren geht.

Integration in Translation-Memory-Systeme

Die Diskussion über den „am besten geeigneten Posteditor“ erübrigt sich aber unserer Meinung nach, wenn die maschinelle Übersetzung im Übersetzungsablauf in Translation-Memory-Systeme (TM-Systeme) integriert wird.

In diesem Fall werden bei der Analyse des zu übersetzenden Textes anhand eines vorhandenen Translation Memory (TM) alle Segmente exportiert, die unterhalb einer vorher definierten Match-Rate liegen (beispielsweise alle Matches unterhalb von 70 %).

