



Abrechnungsmodelle für CAT-unterstützte Übersetzungen

„Houston, wir haben einen Match!“

Zunehmend möchten Kunden von der Arbeits- und Zeitersparnis aus dem Einsatz von CAT-Tools profitieren. Dipl.-Ing. Jerzy Czopik stellt ein Abrechnungsmodell auf Basis gewichteter Wortzahlen vor, das faire Preise für Übersetzer und Kunden ermöglicht – und letztlich sogar als Werkzeug zur Kundenbindung dienen kann.

Bevor wir über das Gewicht nachdenken, muss zunächst Klarheit über die Abrechnungsbasis geschaffen werden. Da die Analysen der Tools meistens wortbasiert sind, brauchen wir einen Wortpreis, den wir der Abrechnung zugrunde legen können. Dieser sollte auf Basis der eigenen Ansprüche bzw. Bedürfnisse errechnet werden. Dazu muss zunächst bestimmt werden, wie viele Wörter eines vollständig neuen Textes man pro Stunde übersetzen kann. Wichtig: Dabei darf nicht der Spitzenwert genommen werden – vielmehr brauchen wir einen Wert, den wir bei jeder Art zu übersetzender Texte dauerhaft erreichen können. Erfahrungsgemäß liegt die realistische Wortzahl eines neuen Textes pro Stunde bei etwa 300. Der Wortpreis für ein neues Wort errechnet sich dann ganz einfach durch die Division Ihres Stundenpreises (d. h. des Stundensatzes, den Sie für sich erzielen wollen bzw. müssen) durch die Anzahl der Wörter pro Stunde. Der Einfachheit halber gehen wir für diesen Artikel von einem Stundensatz von 30 Talern aus, daraus ergibt sich ein Wortpreis von 0,10 Talern. Hierbei rechnen wir nach Ausgangstext ab, was nicht zu vernachlässigen ist.

Der Gewichtungsansatz

Um zu einem Berechnungsansatz zu kommen, werden die Match-Kategorien in verschiedene Klassen eingeteilt und diesen Klassen unterschiedliche Gewichtungsfaktoren zugewiesen. Mit Hilfe dieser Gewichtungsfaktoren erfolgt dann die Korrektur des Preises (nach unten, versteht sich). Diese Vorgehensweise ist mittlerweile für sehr viele Agenturkunden in Deutschland und weltweit üblich, so dass wir uns dieser Tatsache stellen müssen.

Gewichtungsfaktoren und deren Bestimmung

Grundlage für eine Abrechnung nach gewichteten Wortzahlen ist die Festlegung sinnvoller Gewichtungen für die unterschiedlichen Match-Kategorien. Abbildung 1 zeigt eine Beispielanalyse aus SDL Trados Studio:

Typ	Segmente	Wörter	Zeichen	Prozent	Platzierbare Elemente	Tags
PerfectMatch	0	0	0	0.00%	0	0
Kontext-Match	94	94	140	0.95%	94	0
Wiederholungen	447	2917	20261	29.61%	381	52
Cross-file Repetitions	0	0	0	0.00%	0	0
100%	58	90	567	0.91%	29	2
95% - 99%	302	343	1775	3.48%	250	9
85% - 94%	15	63	446	0.64%	6	0
75% - 84%	64	267	1923	2.71%	83	14
50% - 74%	345	2357	17299	23.92%	386	96
Neu	297	3722	26856	37.78%	488	203
Gesamt	1622	9853	69267	100%	1717	376

No Match

Die Gewichtung bei „No Match“ ist einfach: Diese Segmente müssen neu übersetzt werden, sind daher auch voll zu berechnen – die Gewichtung beträgt 100 %. Interessanter ist dagegen, ab welchem Fuzzy-Match-Wert die „No Matches“ beginnen. Die Kategorie „Neu“ entspricht zu 100 % der Definition von „No Match“ – dies sind immer die neuen Wörter. Einige Kunden versuchen jedoch tatsächlich, auch die unterste Match-Kategorie „50%–74%“ mit einer Gewichtung zu versehen. Dies bedeutet: Sogar bei lediglich 51% Übereinstimmung des aktuellen Satzes mit dem im TM abgelegten Segment würde keine volle Wortanzahl abgerechnet (oder nicht der volle Wortpreis bezahlt), sondern eine Minderung stattfinden. Dies ist absolut inakzeptabel, denn bei so niedriger Ähnlichkeit muss der Satz in aller Regel vollständig neu übersetzt wer-



den. Damit ist auch für diese Kategorie die Gewichtung auf 100 % festzulegen. Ähnliche Überlegungen gelten für die Kategorie „75 %–84 %“. Die mögliche Gewichtung bei dieser Kategorie hängt sehr stark von dem verwendeten Tool und natürlich von eigenen Präferenzen ab. Eine Gewichtung dieser Kategorie mit ebenfalls 100 % ist sicherlich kein Fehler.

100 %-Match, Kontext-Match und PerfectMatch

Bei diesen Treffern handelt es sich um Segmente, die direkt aus dem TM übernommen werden können. PerfectMatch stellt hier eine Sonderkategorie dar, denn diese Treffer stehen vorrangig in einem früher übersetzten Dokument, welches in bilingualer und geprüfter (und damit vom Kunden genehmigter) Form vorliegt. Das TM wird hier lediglich zur Segmentierung herangezogen, liefert aber keine Übersetzungen, denn der PerfectMatch-Prozess greift bei der automatischen Vor-Übersetzung des Dokuments nicht auf die Matches aus dem TM zu, sondern nutzt auf Basis des Dokumentvergleiches die im Vorgängerdokument vorhandenen Übersetzungen. Weil dieser Prozess weitgehend automatisch verläuft, ist der Arbeitsaufwand für den Übersetzer an dieser Stelle relativ gering. Da bei Aufträgen von Agenturen in der Regel die Vor-Übersetzung mit PerfectMatch nicht vom Übersetzer selbst, sondern vom Kunden durchgeführt wird, gehe ich in weiteren Überlegungen nicht mehr auf diesen Sonderfall ein.

Die übrigen Arten der exakten Übereinstimmungen bedeuten für den Übersetzer unter bestimmten Umständen tatsächlich verringerten Arbeitsaufwand. Ist die Qualität des TM bekannt, können diese Segmente in der Regel nach einer Überprüfung ohne weitere Bearbeitung übernommen werden. Damit kommt die Bearbeitung dieser Segmente praktisch dem Korrekturlesen gleich. Kennt man nun die Geschwindigkeit in Wörtern pro

Stunde, mit der man im Korrekturmodus arbeitet, ist die Bestimmung der Gewichtung recht einfach. In der Regel kann man davon ausgehen, dass Korrekturlesen mit etwa der dreifachen Geschwindigkeit des Übersetzens erfolgen kann; für unsere Überlegungen nehmen wir 1000 Wörter pro Stunde an. Aus dem Stundenpreis von 30 Talern ergibt sich nun ein Wortpreis von 0,03 Talern – oder eine Gewichtung von 30 %.



Foto: photos.com

Mehr dazu auch bei der BDÜ-Fachkonferenz in Berlin:

Lemster, Czopik et al. – Podiumsdiskussion:
Was kostet ein „Match“ – Sinnvolle Gewichtung oder verdecktes Preisdumping?

So., 30. 9. 2012, 9–10 Uhr, Block 9.4, Hörsaal C

Konferenz- und Tagestickets auf www.uebersetzen-in-die-zukunft.de



Wiederholungen

Wiederholungen sind die Texte, die zwar nicht im TM enthalten sind (und damit beim ersten Vorkommen als „No Match“ gelten), aber in identischer Form im selben Dokument vorkommen. Bei einigen Sätzen bedeutet dies, dass der einmal in diesem Dokument übersetzte Satz problemlos und unverändert an anderen Stellen des Dokumentes eingesetzt werden kann. Der damit verbundene Arbeitsaufwand ist sehr gering, so dass hier sogar eine Gewichtung von nur 10 % denkbar wäre. Andererseits ist es nicht immer so, dass ein Satz ohne weiteres als Wiederholung unverändert übernommen werden kann, wie das folgende Beispiel zeigt:

Ausgangstext	Zieltext
You see a window.	Du siehst ein Fenster.
Click it to select it.	Klicke es an, um es auszuwählen.
You see a file.	Du siehst eine Datei.
Click it to select it.	Klicke sie an, um sie auszuwählen.

Zwar ist das rot markierte Segment eindeutig eine Wiederholung, der Satz kann aber bei seinem zweiten Vorkommen nicht unverändert übernommen werden. In diesem Fall wäre der Arbeitsaufwand größer und damit die Gewichtung von 10 % eher hinfällig. Die praktischen Erfahrungswerte aus dem Markt reichen aber von 0 % (sic!) bis 30 % für Wiederholungen. Je nach Textart und Kunde muss also ein besonderes Augenmerk auf diese Kategorie gelegt werden.

95 %- bis 99 %-Match

Bei dieser Match-Kategorie handelt es sich sehr oft um exakte Übereinstimmungen, die nur aufgrund unterschiedlicher Formatierung (Tagdifferenzen) nicht mehr einen 100 %-Match ergeben. Daraus folgt auch der Gedanke, für diese Match-Kategorie dieselbe Gewichtung wie für die 100 %-Matches zu verwenden.

Übrige Match-Kategorien

Was nun geblieben ist, ist der Bereich zwischen „No Match“ und den hohen Match-Werten. Hier ist eine Begründung der Gewichtung äußerst schwierig, da die Vergleichswerte