

# 22W 520.002 Allgemein: Translationswissenschaftliches Proseminar I AA1

Andreas Hofer

6. Dezember 2022

## 1 Was ist Qualitätssicherung?

Qualitätssicherung bezeichnet die Überprüfung maschinell übersetzter Texte mittels menschlicher Unterstützung, in beispielsweise Post- oder Preprocessing, oder maschineller Unterstützung mittels statistischer Verfahren. Bei menschlicher Qualitätssicherung wird der maschinell übersetzte Text auf grammatikalische oder semantische Fehler überprüft und verbessert oder vor der Übersetzung so aufbereitet, dass bei der automatischen Übersetzung ein besseres Ergebnis erzielt wird. Maschinelle Qualitätssicherung versucht jedoch stattdessen nicht den Text direkt zu verbessern, sondern diesen rein statistisch auszuwerten um so die Qualität der Übersetzung, dessen Güte, zu ermitteln. Diese Bewertung kann schließlich verwendet werden um verschiedene Implementationen der maschinellen Übersetzung zu vergleichen. Einige Beispiele für solche Implementationen sind BLEU (**B**ilingual **E**valuation **U**nderstudy)(Papineni/Roukos/Ward/Zhu 2002), NIST (Von dem US-Amerikanischen **N**ational **I**nstitute of **S**tandards and **T**echnology) (Doddington 2003) und METEOR (**M**etric for **E**valuation of **T**ranslation with **E**xplicit **O**Rdering) (Bannerjee/Lavie 2005), wobei sich BLEU 2002 als erstes Verfahren mit relativ zuverlässigen Ergebnissen behaupten konnte und die meisten anderen gängigen Verfahren auf BLEU aufbauen beziehungsweise dieses weiterentwickelt haben. Doch um die Funktionsweise von Verfahren wie BLEU beschreiben zu können, muss zuerst auf ein paar Grundbegriffe des *Natural Language Processing* (NLP) eingegangen werden.

### 1.1 N-Gramme

N-Gramme sind die grundlegendsten Bausteine, welche in maschinellen Qualitätssicherungsverfahren betrachtet werden. Da der Begriff selbst interdisziplinär verwendet wird, kann ein Baustein eine Vielzahl an Formen annehmen, wobei man in der

Linguistik jedoch oft von einem Buchstaben beziehungsweise einem Wort spricht. Abhängig davon, wie viele dieser Bausteine man miteinander verbindet werden N-Grammen unterschiedliche lateinische Zählpräfixe anstelle des *Ns* angefügt. So besteht ein Monogramm aus nur einem Baustein, ein Bigramm aus zwei usw. Während man theoretisch beliebig viele Bausteine aneinanderhängen kann, sind Mono- bis Tetragramme mit vier Bausteinen am weitesten verbreitet. Wenn man nun mindestens ein Bigramm verwendet, und so jede aufeinanderfolgende Buchstabenfolge innerhalb eines Textes analysiert, kann man statistisch feststellen, mit welcher Wahrscheinlichkeit ein Buchstabe auf einen anderen folgt und auch welche Buchstabenkombinationen nicht, oder nur selten, möglich sind. Dies funktioniert im gleichen Maß mit Wörtern, wobei auch Kollokationen und Redewendungen analysiert werden können.

## 1.2 Genauigkeit und Trefferquote

Genauigkeit und Trefferquote, im Englischen *precision and recall* genannt, geben an in welchem Ausmaß die richtigen Kandidaten gewählt wurden. Genauigkeit definiert die Anzahl der korrekt gewählten Elemente innerhalb der gewählten Elemente, während die Trefferquote die korrekt gewählten Elemente innerhalb aller möglichen korrekt wählbaren Elemente darstellt. Diese beiden Metriken werden oft kombiniert, da eine allein selten die Realität widerspiegelt. Angenommen eine Bilderkennungssoftware soll innerhalb eines Bildes 10 Hunde und 12 Katzen alle Hunde erkennen, dann hätte sie eine hohe Genauigkeit, wenn sie nur Hunde wählt, obwohl sie zum Beispiel nur 2 gefunden hat. Wenn man dies jedoch mit der Trefferquote kombiniert, erkennt man, dass die Software zwar nur richtig gelegen ist, aber auch selten die richtigen Elemente gefunden hat. (Powers 2011:2f)

## Literaturverzeichnis

- Banerjee, Satanjeev/Lavie, Alon (2005) „METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments“, in: *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization* 1:1, 65-72
- Doddington, George (2002) „Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics“, in: *HLT '02: Proceedings of the second international conference on Human Language Technology Research* 2:1, 138-145
- Papineni, Kishore/Roukos, Salim/Ward, Todd/Zhu, Wei-Jing (2002) „BLEU: a Method for Automatic Evaluation of Machine Translation“, in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 40:1, 311-318
- Powers, M. W. David (2011) „Evaluation: From Precision, Recall and F-Factorto ROC, Informedness, Markedness & Correlation“, in: *Journal of Machine Learning Technologies* 2:1, 37-63