• Ihr wollt überprüfen, ob euer neues MU-System ein spezielles linguistisches Phänomen besser übersetzt als ein Standard-MU-System.

Ein spezielles linguistisches Phänomen hat wahrscheinlich nicht besonders viele Testsätze, die man untersuchen kann. Folglich ist es einfacher, die Evaluation manuell vorzunehmen. Auch weil ein so spezifisches Phänomen ein wenig mehr Feinfühligkeit verlangt, als eine Maschine zu bieten hat (Kontrolle ob es richtig ist ist schwieriger).

• Ihr habt ein MU-System für eine Firma gebaut und müsst die Projektleiter über die Qualität des resultierenden Systems informieren.

Hier würden wir eine automatische Evaluation verwenden. Dies, da für eine aussagekräftige Präsentation genug Testsätze ausgewertet werden müssen. Und hierfür einen Menschen anzustellen, kostet sehr viel Geld. Man könnte hier aber auch eine halbautomatische Evaluation verwenden, wenn die Firma genug Geld hat.

• Ihr organisiert einen Wettbewerb für Maschinelle Übersetzung und müsst viele Systeme vergleichen, um den Gewinner festzulegen.

Auch hier wieder die automatische Evaluation. Es lohnt sich einfach nicht so viel Zeit und Geld in diese Auswertung zu stecken. Auch die Menge spielt hier eine Rolle. Wären es nur drei Teilnehmer mit je 2 Testsätzen, könnte man es auch gut von Hand machen.

• Ihr startet ein neues Projekt und möchtet zuerst einen Überblick über die Qualität verschiedener Systeme erhalten, um zu entscheiden, welches davon ihr weiter entwickeln wollt.

Hier die manuelle Evaluation. Denn je nach Thematik des Projekts hat mein eine spezifische Domäne der Testsätze und es ist halt immer noch am genausten, dies von Hand zu bewerten. Ausserdem werden auch nicht tausende Sätze getestet.