# Übung 1

## Aufgabe 1: Google-Übersetzer

Schreiben Sie 10 kurze deutsche Sätze, die der Google-Übersetzer ins Englische übersetzen soll.

Die ersten fünf Sätze sollten Sätze sein, die Google **korrekt** übersetzt.

Die zweiten 5 Sätze sollten ähnlich zu den ersten 5 Sätzen sein, außer dass Google bei der Übersetzung **Fehler** macht. Sie sollten Sätze nehmen, bei denen der Übersetzer alle Wörter kennt (also keine exotischen Wörter oder Namen verwenden). Sie sollten versuchen, Sätze mit unterschiedlichen Typen von Fehlern zu finden.

Speichern Sie die 10 Sätze und die 10 Übersetzungen in eine Datei.

Analysieren Sie, welche Fehler bei den fünf fehlerhaft übersetzten Sätzen vorliegen, und überlegen Sie, wodurch die Fehler verursacht worden sein könnten (bspw. falsche Wortbedeutung übersetzt, falsche Wortstellung, falsche Morphologie). Schreiben Sie zu jedem Satz Ihre Überlegungen dazu. Seien Sie genau bei der Beschreibung und sagen Sie jeweils, von welchen Quellwörtern und Zielwörtern Sie sprechen.

Nun korrigieren Sie die 5 fehlerhaften Übersetzungen **im Google Interface**. Wenn Sie auf verschiedene Teile der englischen Ausgabe klicken, werden Sie sehen, dass Sie Verbesserungsvorschläge für die Übersetzungen bekommen. Sie können auch einfach direkt die Ausgabe im Fenster korrigieren.

Hintergrund: Google speichert diese Verbesserungsvorschläge und benutzt sie, um den Übersetzer zu verbessern. Wir werden in der nächsten Übung testen, ob sich etwas an den Übersetzungen verändert hat. Speichern Sie daher auch die originalen fehlerhaften Google-Übersetzungen für Vergleichszwecke.

Erzeugen Sie zwei Textdateien “german.txt” und “english.txt”: In “german.txt” speichern Sie Ihre 10 Quellsätze (je einer pro Zeile, Satzzeichen etc. durch Leerzeichen abgetrennt). In “english.txt” speichern Sie die 10 Übersetzungen (ebenfalls 1 tokenisierter Satz pro Zeile, die letzten 5 von Ihnen korrigiert).

## Aufgabe 2: Manuelle Wortalignierung

Laden Sie Alex Frasers Wortalignierungseditor herunter:

`http://www.cis.uni-muenchen.de/~fraser/nepal/align_browser_and_german_short.zip`

Paket Sie die zip-Datei aus und lesen Sie die README-Datei

Geben Sie dann folgenden Befehl in der Kommandozeile ein:

```
java TestAlign8
```

Sie sollten nun einen englischen Satz, einen deutschen Satz und eine Alignierung sehen.

Schauen Sie sich die Alignierungen der ersten 20 Sätze an, um ein Gefühl für die Regeln zu bekommen, nach denen hier annotiert wurde. Schauen Sie sich vor allem die komplizierteren Alignierungen an, die keine einfache 1:1-Struktur haben.

Beenden Sie das Programm und führen Sie folgende Schritte aus (die auch in der README-Datei beschrieben sind):

```
rm *.out
```

Hier wird die Ausgabedatei gelöscht. (Wenn Sie bereits Sätze annotiert haben, dürfen Sie diese Datei nicht mehr löschen! Sonst sind Ihre Annotationen verloren!)

```
cp /dev/null align
```

Damit wird die Datei *align* geleert. Sie können die Datei auch mit einem Editor leeren.

```
cp german.txt f
```

Hier kopieren Sie Ihre Datei mit deutschen Sätzen aus der ersten Teilaufgabe nach "f" (nicht "f.txt"!)

```
cp english.txt e
java TestAlign8
```

Jetzt sollten Sie den ersten parallelen Satz sehen. Alignieren Sie ihn durch Clicks mit der linken Maustaste. Wenn Sie fertig sind, klicken Sie auf "next sentence". Annotatieren Sie alle 10 parallelen Sätze.

Dann beenden Sie das Programm und führen folgenden Befehl aus:

```
cp align.out align
```

Damit werden die annotierten Alignments dauerhaft gespeichert!

```
rm *.out
```

Dieser Befehl erlaubt Ihnen ggf. die Alignierung neu zu starten.

Fassen Sie schriftlich zusammen, welche Wörter schwierig zu alignieren waren (bspw. englische Funktionswörter ohne klare Entsprechung auf der deutschen Seite) Begründen

Sie Ihre Entscheidungen bei schwierigen Fällen.