

Übung 2

Aufgabe 1: Google-Übersetzer Revisited

Nehmen Sie Ihre fünf Sätze aus der Übung 1, die Google fehlerhaft übersetzt hat und lassen Sie sie erneut übersetzen.

Bekommen Sie immer noch dieselbe Übersetzung oder wurden Ihre Verbesserungsvorschläge teilweise oder vollständig übernommen?

Aufgabe 2: Computerunterstützte Übersetzung mit OmegaT

Laden Sie OmegaT von der Seite omegat.org herunter.

Erzeugen Sie ein neues Projekt (Lesen Sie dazu die “In 5 Minuten zur Nutzung von OmegaT!”-Anleitung unter <https://omegat.sourceforge.io/manual-standard/de/index.html>). Nennen Sie das Projekt *mytest*. Wählen Sie als Quellsprache EN-US oder EN-GB und als Zielsprache DE-DE. Notieren Sie sich, wo das Projekt erzeugt wurde (den Verzeichnispfad).

Wechseln Sie ins Hauptverzeichnis des Projektes und dann in das *source*-Unterverzeichnis und erzeugen Sie eine Textdatei namens *text1.txt* mit 5 englischen Sätzen (bspw. Sätzen aus der letzten Übung). Verwenden Sie korrekte Interpunktion. OmegaT kann die Sätze selbst segmentieren.

Starten Sie OmegaT und laden Sie das Projekt. Sie sollten 5 Sätze sehen, die für die Übersetzung anstehen. Klicken Sie auf den Zielteil von jedem, und geben Sie die deutsche Übersetzung ein.

Wählen Sie “generate translations” (Tastenkürzel Strg-D), damit OmegaT seine Datenbank von Übersetzungen im “target”-Unterverzeichnis speichert.

Speichern Sie und verlassen Sie OmegaT.

Die Ergebnisse Ihrer Arbeit sind in dem “target”-Unterverzeichnis unter demselben Dateinamen gespeichert. Prüfen Sie, ob der Inhalt korrekt ist.

Wechseln Sie dann ins “source”-Verzeichnis des Projektes und erzeugen Sie eine weitere Textdatei “text2.txt”. Als ersten Satz nehmen Sie den ersten englischen Satz aus “text1.txt”. Fügen Sie drei neue Sätze hinzu. Diese sollten ähnlich zu den Sätzen 2 bis 4 aus “text1.txt” sein. Ändern Sie nur ein Wort pro Satz.

Starten Sie OmegaT und laden Sie das Projekt. Sie sollten 4 Sätze sehen. Der erste Satz sollte genau gleich sein. Akzeptieren Sie ihn. Dann klicken Sie auf den zweiten Satz. Sie sollten jetzt einen “fuzzy match” auf der rechten Seite sehen. Drücken Sie die rechte Maustaste, um zu “Replace translation with match” zu gelangen. Dann editieren Sie. Editieren Sie alle Sätze.

Wählen Sie “generate translations” (Tastenkürzel Strg-D), damit OmegaT seine Datenbasis mit Übersetzungen im “target”-Unterverzeichnis speichert.

Speichern Sie und beenden Sie OmegaT.

Schauen Sie sich nun den Inhalt der Datei “mytest-omegat.tmx” im Hauptverzeichnis des Projektes an und diskutieren Sie den Inhalt. Wozu dient diese Datei? Wie sollten Sie sie verändern, um die Übersetzungsrichtung zu ändern (also Deutsch nach Englisch zu übersetzen)? Wieviel Unterstützung für die Segmentierung und das Fuzzy Matching im Deutschen gibt es (siehe das OmegaT-Handbuch)? Vergleichen Sie das mit der Unterstützung der Segmentierung und des Fuzzy Matchings in Englisch.

Sprachmodelle, IBM-Modell 1

Berechnen Sie die Parameter $p(e|e')$ eines Bigramm-Sprachmodelles für das Trainingskorpus mit den folgenden vier Sätzen:

START the house is small END
 START the house is little END
 START the building is small END
 START the home building is small END

Berechnen Sie für den Satz \mathbf{f} = “das Haus ist klein” und die folgenden Übersetzungswahrscheinlichkeiten $p(f|e)$

the		house		building		is		exists		little		small	
der	0.3	Haus	0.5	Gebäude	0.7	ist	0.7	ist	0.2	klein	0.7	klein	0.8
die	0.3	Heim	0.1	Haus	0.3	existiert	0.2	existiert	0.7	kurz	0.3	kurz	0.2
das	0.4	Gebäude	0.4			hat	0.1	hat	0.1				

die Wahrscheinlichkeit $p(\mathbf{e}, \mathbf{f})$ der folgenden Übersetzungen

\mathbf{e} = “the building is little”
 \mathbf{e} = “the house exists small”

auf Basis der oben berechneten Sprachmodellparameter mit IBM Modell 1:

$$p(\mathbf{e}, \mathbf{f}) = p(\mathbf{e}) p(\mathbf{f}|\mathbf{e})$$

$$p(\mathbf{f}|\mathbf{e}) = \frac{\varepsilon}{(|\mathbf{e}| + 1)^{|\mathbf{f}|}} \prod_{i=1}^{|\mathbf{f}|} p(f_i|e_{a_i})$$

Sie können dabei die Konstante ε ignorieren.