

## 1. Übungsblatt - Abgabe: 29.10.2013

### Aufgabe 1.1 Maschinelle Übersetzung

- (a) Auf den Webseiten der Computerlinguistik findet sich unter <http://www.coli.uni-saarland.de/msc/departement.html> ein Absatz mit der Überschrift “Computational Linguistics Department in Saarbrücken”. Kopieren Sie diesen Absatz (bis zu “... multimodal interfaces”) in das Textfenster des google translators (unter <http://translate.google.com/>), und lassen Sie ihn ins Deutsche übersetzen. Wie würden Sie Qualität und Stil der Übersetzung generell kommentieren? (1-2 Sätze)
- (b) Korrigieren Sie den übersetzten Text: Markieren Sie fehlerhafte oder stilistisch schlechte Übersetzungen, geben Sie an, welche Art von Fehler gemacht wurde (ähnlich, wie Sie das aus der Schule noch kennen: „Ausdruck“, „Satzbau“ etc.), und geben Sie bessere Übersetzungsalternativen an (für die beanstandeten Stellen oder für ganze Sätze).
- (c) Probieren Sie das System auch mit ein paar (mindestens 2) englischen Texten von anderen Textsorten aus: News, Gedicht, persönliche Email. Wie ist das Ergebnis hier im Vergleich? Geben Sie Originaltext und Übersetzung an begründen Sie Ihre Einschätzung mit Beispielen.

### Aufgabe 1.2 Dialogsysteme – Ersatz

- (a) Rufen Sie das Bahnauskunftssystem der Deutschen Bahn an (Tel. 0800/1507090), und testen Sie es, indem Sie einige Zugverbindungen erfragen. Versuchen Sie, das System durch Ihre Fragen auszutesten und an seine Grenzen zu bringen. Berichten Sie! Wo hatten Sie mit dem System Probleme? Wo hatte das System besondere Probleme?
- (b) Versuchen Sie, möglichst kurz zu beschreiben, wie die Bahnauskunft Ihrer Meinung nach funktioniert: Was versteht sie? Was weiß sie? Was kann sie?
- (c) Versuchen Sie, das Muster, das der Dialogführung der Bahnauskunfts zu Grunde liegt, möglichst übersichtlich aufzuschreiben (nicht unbedingt vollständig): Wie reagiert das System in bestimmten Situationen auf bestimmte Benutzeräußerungen? Wie sieht dementsprechend der „Dialogplan“ der Bahnauskunft aus?

### Aufgabe 1.3 Informationssuche

- (a) Benutzen Sie Google ([www.google.com](http://www.google.com)), um die Antwort auf die Frage „Wer war US-Präsident als Barack Obama geboren wurde?“ bzw. “Who was American president when Barack Obama was born?” zu erhalten. Beschreiben Sie, wie Sie vorgegangen sind, und die Schwierigkeiten, die man hat, wenn man die Frage durch eine einzige Google-Anfrage beantworten möchte.
- (b) Finden Sie drei weitere Fragen, deren Antworten aus möglichst unterschiedlichen Gründen nur schwer oder gar nicht mit Google zu finden sind (möglichst nicht nur, weil man mehrere Suchanfragen miteinander kombinieren muss), deren Antworten aber grundsätzlich im Internet vorkommen. Erläutern Sie, warum diese Fragen jeweils schwierig sind.

### Aufgabe 1.4. Ambiguität

Lesen Sie aus Jurafsky/Martin (Neu-Ausgabe von 2009) Kapitel 1, S. 35-43, Abschnitte 1.1, 1.2, 1.4 und 1.5. Die in 1.3 verwendeten Begriffe sind nicht ohne weiteres verständlich, sie werden im Lauf der Vorlesung eingeführt; 1.3 kann übersprungen werden. Der Text steht in Kopie in der Institutsbibliothek (Vorlesungsordner).

- (a) Versuchen Sie, für die 5 Lesarten des Satzes “I made her duck” aus 1.2 möglichst eindeutige und gut verständliche deutsche Formulierungen zu finden.
  - (b) Versuchen Sie, einen deutschen Satz mit maximal 4 Wörtern zu finden, der möglichst viele, aber mindestens 2 Lesarten besitzt. Paraphrasieren Sie die einzelnen Lesarten und erklären Sie, wodurch die Ambiguität zustande kommt.
- 

### Feedback

Bitte beantworten Sie kurz folgende Fragen:

- (a) Wie schwer fanden Sie den Stoff der aktuellen Vorlesung? (sehr einfach, einfach, durchschnittlich, schwer, sehr schwer)
  - (b) Wie viel Zeit haben Sie mit der Bearbeitung des aktuellen Übungsblattes verbracht?
  - (c) Haben Sie sonstige Fragen oder Anmerkungen zur aktuellen Vorlesung? Wenn ja, welche?
- 

Abgabe in Gruppen von bis zu drei Studierenden bis **29.10.2013** 10 Uhr entweder als Email- Anhang im pdf-Format an **[i2cl@coli.uni-sb.de](mailto:i2cl@coli.uni-sb.de)** mit dem Betreff **i2cl homework 1** und Ihrer Matrikelnummer als Dateinamen oder auf Papier im Briefkasten an der Tür von Raum 1.04 in C7.2.