

**Übungsklausur (Gesamtpunktzahl: 120 Punkte)**

1. Gegeben sei eine kontextfreie Grammatik mit den Produktionsregeln

|    |           |     |   |
|----|-----------|-----|---|
| S  | -> NP VP  | V   | -> spielte, gähnte                          |
| VP | -> V NP   | N   | -> Musiker, Sinfonie, Paukenschlag, Klavier |
| VP | -> V      | Prp | -> auf, mit                                 |
| VP | -> VP PP  | Det | -> der, dem, den, die, der                  |
| NP | -> NP PP  |     |   |
| PP | -> Prp NP |     |   |
| NP | -> Det N  |     |   |

Leiten Sie den Satz *Der Musiker spielte die Sinfonie mit dem Paukenschlag* auf zwei unterschiedliche Weisen ab. Geben Sie die beiden Strukturbäume an, die aus der Ableitung resultieren. (15 Punkte)

2. Leiten Sie mit der Grammatik aus 1. zwei Wortfolgen ab, die aus unterschiedlichen Gründen keine korrekten Sätze des Deutschen sind. Kommentieren Sie, gegen welche unterschiedlichen grammatischen Regeln dabei verstoßen wird. (10 Punkte)

3. Nennen Sie die drei in der Vorlesung eingeführten Motivationen, Computerlinguistik zu betreiben, und charakterisieren Sie jeweils kurz, worin das spezifische Ziel dieser Motivationen liegt. (5 Punkte)

4. Auf die Frage „Which European Country has the largest population?“ liefert das QA-System Answerbus (<http://answerbus.coli.uni-saarland.de/>) die drei folgenden Ergebnisse. Kommentieren Sie die unterschiedlichen Antworten in Bezug auf Relevanz (ist die Antwort informativ für die Frage?) und Korrektheit (beantwortet das Dokument tatsächlich die gestellte Frage?). (10 Punkte)

1. [wikianswers.com] „What country in the European Union has the largest population? – Germany with about 82.5 million.“
2. [allbusiness.com] „China, the largest population country in the world, approves a patent for cancer diagnosis that is still pending in most European countries.“
3. [wikianswers.com] „Which country in Europe has the largest population? – Germany has the biggest population in Europe.“

5. Nennen Sie drei Herausforderungen der Text-to-Speech-Sprachsynthese, und beschreiben Sie sie kurz. (10 Punkte)

6. Sie haben zu Beginn der Vorlesung die beiden "klassischen" Sprachverarbeitungssysteme ELIZA und SHRDLU kennengelernt. Der wesentliche Unterschied zwischen beiden

Systemen ist, dass SHRDLU im Gegensatz zu ELIZA wissensbasiert arbeitet. Erläutern Sie diese Feststellung an je einem Beispiel (d.h. einer Dialogsequenz mit System- und Benutzeräußerung) zu den beiden Systemen. Sie können die Beispiele frei, und falls Ihnen das lieber ist, auch in deutscher Sprache, wählen. (10 Punkte)

7. Eine naive Methode, um im Internet Dokumente zu identifizieren, die für eine Anfrage relevant sind, ist zu testen, ob Worte aus der Anfrage im Dokument vorkommen. Dieses naive Vorgehen ist sowohl unvollständig (es findet nicht alle relevanten Dokumente) als auch inkorrekt (es findet irrelevante Dokumente). Beschreiben Sie für beide Fehlerarten kurz je ein linguistisches Phänomen, das zu dieser Fehlerart führen kann, und geben Sie ein konkretes Beispiel. (10 Punkte)
8. Die Potenzautomatenkonstruktion erzeugt aus einem buchstabierenden NEA mit  $n$  Zuständen einen DEA mit  $2^n$  Zuständen. Der Potenzautomat für einen NEA mit 10 Zuständen hat also 1024 Zustände. Trotzdem kommt man beim Aufschreiben des DEA meist mit einer weit kleineren Zustandsmenge aus. Wieso? (10 Punkte)
9. Geben Sie einen endlichen Automaten an, der die Sprache  $L$  über dem Alphabet  $\{a,b\}$  mit  $L = \{w \mid w \text{ enthält höchstens ein } a \text{ und mindestens ein } b\}$  beschreibt. (15 Punkte)
10. Nennen Sie die Arbeitsschritte, die zur Erstellung eines syntaktisch annotierten Korpus (einer „Baumbank“) nötig sind. Beschreiben Sie sie kurz. (10 Punkte)
11. Gegeben sei der folgende NEA:  
 $A = \langle \{1,2,3\}, \{a,b\}, \Delta, 1, \{2,3\} \rangle$  mit  
 $\Delta = \{ \langle 1, ba, 1 \rangle, \langle 1, b, 2 \rangle, \langle 1, b, 3 \rangle, \langle 2, bab, 3 \rangle, \langle 3, \epsilon, 2 \rangle \}$   
Konstruieren Sie nach dem in der Vorlesung eingeführten dreistufigen Verfahren einen zu  $A$  äquivalenten DEA! (15 Punkte)