

## Übungsklausur

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.

- 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.
- Erklären Sie kurz, was Rauschen (Noise) in einem Korpus ist, und weshalb es ein Problem darstellt. Geben Sie ein konkretes Beispiel.
- Eine probabilistische kontextfreie Grammatik weist den syntaktischen Analysen eines Satzes Wahrscheinlichkeiten zu. Dadurch lassen sich die Analysen ordnen.
  - Erklären Sie kurz, wieso und für welche Zwecke diese Ordnung wichtig ist.
  - Schildern Sie kurz einen linguistischen Zusammenhang, der durch Wahrscheinlichkeiten der Form  $P(NT \rightarrow RS)$  nicht modelliert werden kann, und geben Sie ein konkretes Beispiel.
- Geben Sie eine kontextfreie Grammatik an, die arithmetische Gleichungen der Form

$x=y,$   
 $(x+y)=(z:x),$   
 $(x*(y-(-z))=(-x)*(-y))$

erzeugt. Das heißt, Terminalsymbole sind  $x, y, z, =, +, -, *, :, ($  und  $)$ , und alle komplexen Terme werden in Klammern geschrieben. Verwenden Sie als nicht-terminale Symbole S (Gleichung), T (Term), O (Operator).

6. In der Vorlesung wurden "Sprachassistentz", "Dialog" und "Informationsmanagement" als drei wichtige Gruppen sprachtechnologischer Anwendungssysteme genannt. Geben Sie für jede der drei Gruppen zwei möglichst unterschiedliche Beispiele an, und zwar nicht die Namen bestimmter Systeme, sondern unterschiedliche Anwendungen: Also z.B. nicht "LIFT", sondern "Dialogschnittstelle zur Geräte- bzw. Fahrstuhlsteuerung".
7. Lexikalische Information in Sprachverarbeitungssystemen kann alternativ in Wortformenlexika und in Stammlexika bereitgestellt werden. Was sind Vorteile und Nachteile? Bitte kurz und stichwortartig beantworten! Wieso ist es für das Deutsche wichtiger als für das Englische, mit einem Stammlexikon zu arbeiten?
8. 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).
9. 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. 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?
11. 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 !
12. 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!