

# Einführung in die statistische Sprachverarbeitung

Detlef Prescher

April 23, 2007

## Hausaufgabe 2

A) Bestimme die empirische Wahrscheinlichkeitsverteilung der Worte und der Punctuation im folgenden Korpus (des Dichters Joachim Ringelnatz):

Man mische 7 Pfund Palmin mit gleichviel Milch und Terpentin.  
Dann füge man ein Hühnerei und etwas Öl nebst Essig bei. Dies  
nun zu festem Brei gerührt, wird dann in einen Strumpf geschnürt.  
Das ganze lässt man 13 Wochen in lauem Seifenwasser kochen. Dann  
wird es mit Gelee garniert und im verdeckten Topf serviert.

B) Berechne die Wahrscheinlichkeit des Korpus mit der Verteilung aus Aufgabenteil A.

C) Fasse nun jeden Satz als einen eigenen Korpus auf und berechne seine Wahrscheinlichkeit mit der Verteilung aus Aufgabenteil A. Welcher Satz ist am wahrscheinlichsten?

D) Gib zu jedem Satz eine weitere Verteilung an, die diesen Satz eine höhere Wahrscheinlichkeit als die in C zuweist. Kannst Du auch eine Verteilung finden, die dem Gesamtkorpus eine höhere Wahrscheinlichkeit als die in B zuweist?

E) Beweise, dass sich die Wahrscheinlichkeiten aller Korpora einer beliebigen, aber festen Länge  $n$  sich zu 1 aufsummieren! Relativiere Deine Aussagen in Aufgabenteil C.

**Abgabetermin: Montag, 30. April**

**Besprechung: Mittwoch, 2. Mai**